NATURE · VOL 366 nature 16 DECEMBER 1993

Intellectual property under GATT

This week's GATT agreement (if signed) leaves doubts about the protection of intellectual property unresolved. With UNESCO absent from the scene, the world's academies should take up the issue.

LAST-MINUTE brinkmanship by international trade negotiators notwithstanding, the latest instalment of the General Agreement on Tariffs and Trade (GATT) is likely to have been signed at Geneva by the time this issue of *Nature* appears. It will have been a close call, and a tiresome one, but there are several levels on which the agreement should be welcomed. One is that, by stimulating an increase of world trade, an agreement will increase the prosperity of member states. More generally and over many years, the agreement will induce a better approximation to an efficient division of labour between the member states; some will grow food, others will make supercomputers, as their skills and inclinations dictate and as the working of a competitive market allow. But the interdependence fostered by mutual dependence on tradeable goods is a useful inhibitor of more open quarrelling, and so should make for a more peaceful world. Who can quarrel with that?

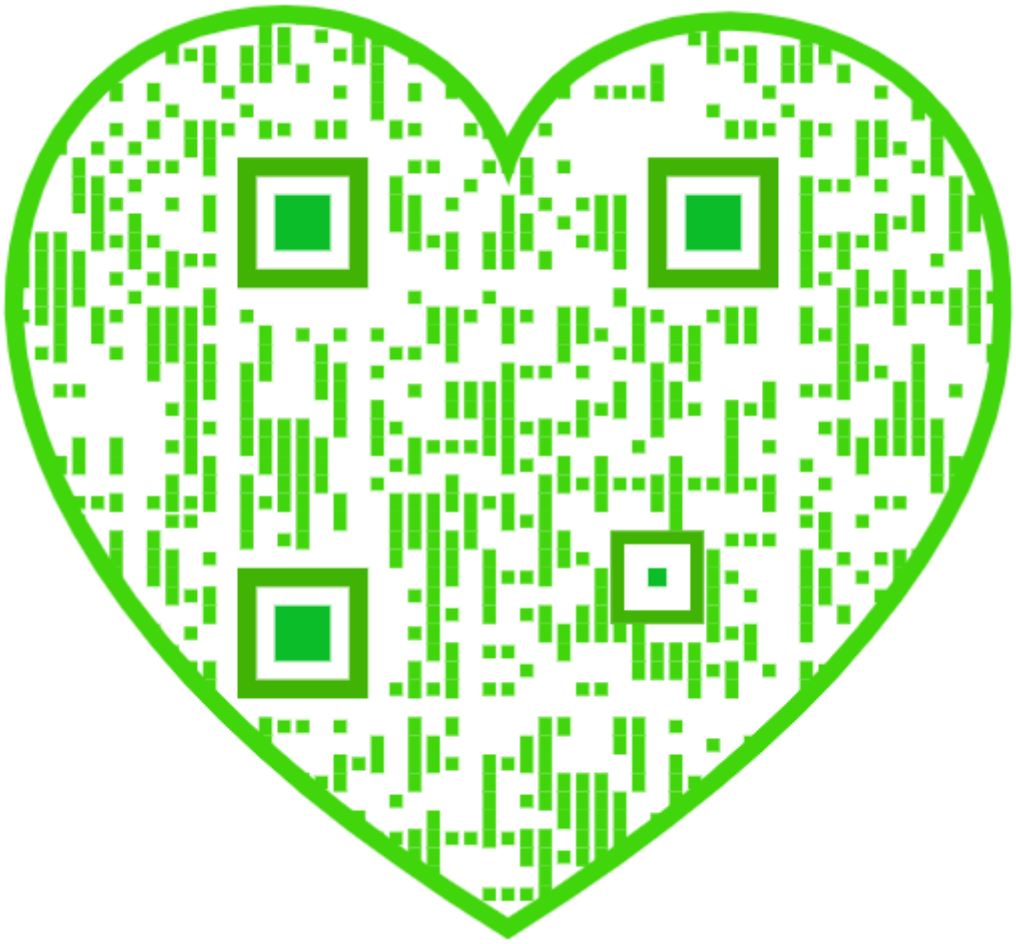
Several people, as the world's newspapers have shown in the past few weeks. But among the objections from out-and-out protectionists (whose minds should be changed when the new agreement comes into effect in 1995) are buried some that need to be taken seriously. One general objection is that the industrialized world remains excessively protective of its textile industries, to the disadvantage of developing countries. The Multi-Fibre Agreement, which allows countries to impose quotas on their imports of textiles. That offends against the spirit of the GATT, which makes quotas anathema. Negotiated (some developing countries say "dictated") separately from the GATT, it should be separately scrapped, and quickly.

Another issue in need of re-examination is the protection of intellectual property, which has become more important in this GATT round because of the partial extension of the agreement to the trade in services. But it has been plain from the outset that an arrangement such as GATT, whose fundamental principle is that countries should not discriminate between potential imports on offer on the world markets, would not be workable without some harmonization of domestic legislation on the protection of intellectual property, chiefly patents and copyrights. Otherwise it would always be possible for, say, mousetrap manufacturers in a country offering no patent protection to undercut the prices charged by manufacturers in countries where new products are protected by patents; at the very least, manufacturers in the former would avoid the payment of a royalty to the original inventor. So there has to be an understanding between GATT members on the kind of patent legislation each of them will have. So much is agreed.

What the present GATT round has not agreed is what kind of legislation that should be. Instead, the agreement says simply that the member states should be free to keep their present legislation in force for a period of four years, during which there will be further negotiations on the subject. Presumably it is intended that a deal should be struck in time for it to be in effect by the end of the century or a little earlier. That invites the kind of horse-dealing that has made the past few weeks at Geneva a kind of circus. The research community, whose inventions are the commodity most at issue, should have a say in what happens in the next few years.

A few principles about the present regime stand out. First, unless inventors enjoy some kind of material reward, there will be fewer inventions (which is why the patenting system was itself invented). Second, a monopoly right on exploitation, but one that lapses after a specified period, is an equitable way of rewarding inventors. Third, an inventor granted a patent who then declines to exploit it will be at the mercy of those who will. Fourth, there may be occasions when a government is so offended that its people are being held to ransom by a rapacious inventor that it is free to licence the patent to whom it chooses, fixing a rate of royalty and paying it over to the original inventor. There is a closely related administrative matter. Securing patent protection is not cost-free, but often so costly that patenting in a representative range of countries can cost as much as what has been spent on research and development in the first place.

The past few years have brought further complications, not least the notion that it should be possible to patent the nucleotide sequences of parts of genes, notably those generated as complementary-DNA (cDNA) from RNA molecules found in particular tissue cells. The first applications by the US National Institutes of Health in respect of some hundreds of such sequences were rejected last August by the US patents office, but NIH say they will appeal against. They have lawyers logic on their side. These cDNA tags are artefacts (which have been made, or are not found naturally), they are useful (in pulling out the genes to which they belong) and it is not "obvious" (in patent-lawyers' language) that all that should be so. The research community has been offended that people should ever have considered applying for patents for entities such as this, and has said so. But if patent protection is denied, the commercial companies with an interest in the field will simply never publish what they



know, but will let their knowledge of the human genome accumulate until they can patent an indubitably artefactual material, a drug perhaps.

The developing countries raise other problems. This week's survey of Science in India draws the attention in another direction. There is no way in which a peasant farmer can be told that he cannot save his seed from one crop to the next, or forbidden by regulations sanctified in a far-off city called Geneva not to sell surplus seed to somebody else. Worse still, there is no way in which a traditional use of the product of an indigenous plant, say the contraceptive ingredient of neem-tree oil, should become taxable when some company overseas discovers the active principle. There would be riots if there were attempts to enforce a doctrine such as that. There would be more modest trouble if improvements of the millet crop engineered in California were charged for. But what incentive will there be for people with the skill to tackle these problems if there is no reward?

That is why the research community needs to take this issue in its grasp. Science is about understanding the world, but it also has the property of creating wealth in which, in equity, it (and its people) should have a share. Everybody agrees. The difficulty is that the now-conventional rules for the protection of intellectual property from which wealth may spring no longer correspond with professional people's estimates of the efforts expended in acquiring it, nor on the needs of the real world (and of the needs of the developing world in particular). Patent lawyers are fond of referring to the antiquity of the system they help administer, which should be a giveaway; it has also outlived its applicability. Ideally, it would be for UNESCO to take up this question, but that organization is not yet sufficiently restored to health to meet the need. The world's academies, which held a successful if unremarkable conference on population growth in New Delhi in October, are better placed to take up the challenge. If they will not, who will? □

Decriminalizing drugs?

The US Surgeon-General has revived a long-simmering debate about existing legal restraints on drug abuse.

JOYCELYN Elders, the Surgeon-General of the United States, last week gave new life to a long-standing argument about the criminal sanctions applied to drug abuse when she said, in off-the-cuff remarks, that "I do feel that we would markedly reduce our crime rate if drugs were legalized." The reaction has been swift and hostile. Conservative Republicans in Congress went wild, many of them calling for her dismissal and saying that this remark was the last straw. (The outspoken Surgeon-General talks frequently and bluntly against teenage pregnancy, especially among poor minorities, and advocates condoms and any other measures that will keep children from having children.) Within hours, the White House also repudiated Elders' comments, saying that President Bill Clinton is opposed to legalizing drugs. "End of story. There is no more to discuss", the White House press secretary said.

But not so fast. Although Elders forthrightly admitted that she is not a scholar in the field of drug decriminalization, she is quite right to suggest that the idea should be given serious study. In the United States and elsewhere, the use of illicit drugs (ranging in potency from marijuana to crack cocaine and heroin) is creating a generation of hopelessly addicted and crime-driven young men (and women) in the worlds' inner cities. In a vicious cycle of addiction and crime, a frightening increase in robbery and murder are tied together by a craving for drugs, a need to steal to buy them and the violent defence of trading pitches by the pushers.

But the connections may not be as simple as they seem. One possibility, tested somewhat in the Netherlands, is that if drugs were legal, crime would be reduced simply because the cost of buying them would decrease. There are also data from the Netherlands, where the use of marijuana is not illegal, to suggest that decriminalization does not bring increased use. To be sure, the data by themselves do not make a compelling case for legalization. But they are better than mere straws in the wind.

The arguments against the legalization of drugs, or some of them, are often different in kind, asserting that legalization would be seen as "giving in" to drug cartels and dealers, and would signal to people that it is safe to get high on crack cocaine. But would it be such a blow to the collective pride if the cartels and dealers were put out of business? And is it necessarily the case that licit use would become excessive and addictive use? Alcohol, legal and socially acceptable in all but *shi'ite* Moslem states, is not an irrelevant example; a little may be good; too much is not only bad for heart and liver, but turns otherwise law-abiding people into public dangers behind the wheel of a car. What evidence is there that marijuana in small amounts is more dangerous?

The more serious argument is that use of drugs inexorably leads to addiction and that the use of relatively harmless marijuana leads to the use of more potent drugs, cocaine, crack and heroin. The argument would be more powerful if there were compelling data to support it, but it is also possible to turn the supposition on its head and to conclude that if marijuana were legalized, but other drugs were not, those with a hankering after artificial psychosensation would satisfy their needs with the cheap and licit substance on the market, to the neglect of the harder stuff (and the chagrin of those who deal in it).

All that, of course, is handwaving. The truth is that there is insufficient understanding of how people and the societies of which they are a part would respond to the legalization of drugs of any potency. What Joycelyn Elders has done is to draw attention to the need to learn more about the role of drugs in modern society. Especially when "prevention" is all the rage, as now in the United States, there is the strongest case for a fresh government supported attempt to understand the social connotations of the use of now illicit drugs. What harm would be done by that? After all, if building prisons, locking people up and interdicting drugs at national borders were the solution, the problem would have been be eradicated by now. □

Delors white paper puts research firmly on Europe's political map

Paris. Europe's political leaders have given their formal endorsement to proposals that would enlarge their common research programmes, encourage the joint construction of new 'information highways' and increase incentives to persuade industry to invest in research.

At their summit meeting in Brussels last weekend, the leaders of the 12 member states of the European Union (EU) approved a white paper (policy document) on competitiveness, growth and unemployment prepared by Jacques Delors, the president of the commission. The paper is aimed at creating 15 million jobs in the EU by 2000.

Three of its ten chapters dealt exclusively with research and development (R&D). The EU's acknowledgement of the importance of R&D to its plans for economic recovery was further reinforced at the summit, when the heads of government also broke the deadlock over the funding of the EU's next five-year Framework programme, by approving a budget of ECU12 billion (US$13.6 billion, see sidebar).

In his white paper, Delors sets the EU the ambitious target of increasing spending on R&D to three per cent of gross national product (GNP). It now spends just 2 per cent (ECU104 billion) in contrast with the United States, which spends 2.8 per cent (ECU124 billion) and Japan three per cent (ECU77 billion).

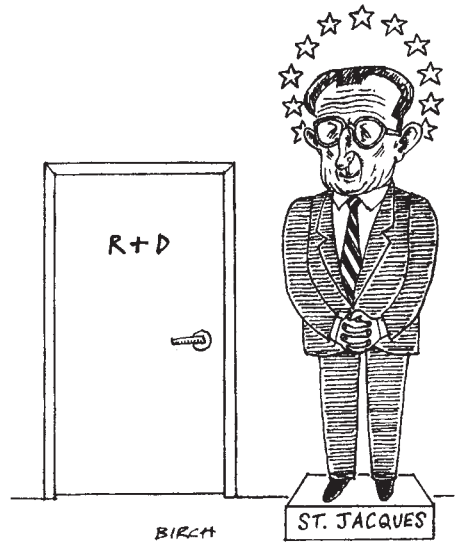
Such a large increase in spending seems improbable, given that member states have either frozen or cut science spending because of the economic downturn. Nevertheless, Delors is confident that the private sector could make up the difference. Companies fund just over half of all science spending in Europe, compared with more than three-quarters in Japan.

To this end, Delors encourages member states to provide tax and other incentives to companies to invest in research. He also wants the EU to make its rules for co-funding industrial research more flexible (see *Nature* **365**, 775; 1993).

Delors also criticizes the lack of coordination between national research policies. His remedy would be to formalize cooperation within some form of European science agency, but it is too soon to say how this would operate. He also wants member states to take joint measures to improve technology transfer.

He recommends that national research organizations, companies and social groups need to develop a European strategy for biotechnology as a matter of urgency. He has also instructed the commission to consider revising legislation affecting biotechnology products.

The white paper proposes that governments encourage companies to work together on several big projects in the fields of information technology, biotechnology and environmental technology. The European Round Table, which brings together 40 leading industrialists, backs the plan.



The paper suggests ECU150 billion should be spent on information technology infrastructure over the next ten years. Furthermore, it accords priority to eight projects, including building a high-speed communications network and developing databases and electronic mail, requiring ECU67 billion in 1994–98.

Delors wants the EU to set up a high-level "Task Force on European Information Infrastructure" to plan the programme and start it by the middle of next year. Although EU has agreed to provide ECU12 billion a year for six years to create networks in transport, energy and telecommunications, it anticipates that most of the money for the electronic highway will come from the private sector.

It is too soon too say what effect the EU summit's adoption of the Delors plan will have on science and technology. Although the member states have committed themselves to implementing the white paper's recommendations, these are non-binding and will inevitably be subject to change.

Declan Butler

Deadlock broken on Framework funding

Paris. The leaders of the European Union (EU) ended uncertainty about the next five-year Framework research programme last Friday by agreeing to a budget of "at least" ECU12 billion (US$13.6 billion), and a reserve of ECU1 billion, for the period 1994–98.

The heads of government acknowledged that the research programme was an important part of its plans for economic recovery (see above). In the final communique, they declared that the implementation of "an ambitious, well-targeted research programme constitutes a significant contribution to efforts towards recovery, particularly in areas such as information technology".

The programme had been hanging in the balance earlier in the week, after the council of research ministers failed to agree on the budget of ECU13.1 billion proposed by the European Commission, and had passed the buck to the summit (see *Nature* **366**, 499; 1993).

The commission is now confident that the council of ministers will iron out its remaining disagreement over the funding of the EU's four Joint Research Centres at an extraordinary meeting a few days before Christmas. The Framework programme is then expected to pass easily though its final reading before the European Parliament, and to come into effect some time next year. **D. B.**

Newspaper ducks criticism of AIDS coverage

London. Under a 360-point heading set in blood-red type "AIDS — why we won't be silenced", Britain's *Sunday Times* this week (12 December) accuses *Nature* of "sinister intent" and of being "so tied to one particular hypothesis of Aids as to make it feel anybody who questions that theory must not only be in error, but must be silenced".

In particular, the newspaper criticizes John Maddox, editor of *Nature*, for his alleged refusal to publish articles "from doctors and scientists who question the HIV theory of Aids". The newspaper describes Maddox as the "high priest" of *Nature*, and the journal as the "bible of the church of science". But it also concedes that the HIV virus may be involved in the onset of AIDS in some unspecified way.

The *Sunday Times* reprints the leading article from last week's *Nature* (**366**, 493–494; 1993), and purports to respond to the questions raised. But the points made ▶

by Neville Hodgkinson, the science correspondent of the *Sunday Times* and the author of the article, are familiar ones, and little attempt is made to provide detailed answers to the responses to claims made in earlier articles in the newspaper.

Much is made of the fact, for example, that new estimates of the global pattern of AIDS spread are not as high as some of the earlier estimates. But, as Professor Roy Anderson pointed out earlier this year (*Nature* **363**, 393–394; 1993), the new figures are within the bounds of the original estimates, which were themselves known at the time to be subject to considerable uncertainty because of the lack of data which have since been collected.

Similarly, Hodgkinson repeats the claim that HIV-positive haemophiliacs have recovered after being treated for a blood-clotting disorder. But this claim is dismissed by the head of a haemophilia centre at one of London's main teaching hospitals, who says that she has explained to Hodgkinson why the view is wrong, and that repeating it is deeply distressing to affected people who may be given false hope.

(The comments on the "HIV theory" with respect to haemophiliacs is particularly odd given that, in a different part of the same edition of the *Sunday Times*, the paper congratulates itself for helping to win compensation for a group of British haemophiliacs who had been infected with HIV from contaminated US blood.)

Despite the aggressive style of the two-page attack on *Nature* and the (unnamed) supporters of what the newspaper describes as the "HIV theory", Hodgkinson appears to be backing off from the newspaper's position only the week before, when it promoted the views of a small but vociferous group of self-styled "dissidents" who do not believe that HIV causes AIDS.

In this week's article, Hodgkinson says "we have never argued that any of these findings rule out a role for HIV in Aids". His main complaint now is that *Nature* has not, he claims, examined the significance of the so-called anti-HIV results.

Such statements appear to ignore the many occasions on which *Nature* has indeed assessed some of the less generally accepted views on AIDS, most recently, for example, by M. Ascher *et al.* in *Nature* **364**, 291–292; 1993.

Once again, the *Sunday Times* repeats the views of a few people who do not work with AIDS patients and say it is a disgrace that *Nature* has rejected their speculative hypotheses. In neither of its articles in the past two weeks has it reported the views of clinical researchers working with AIDS patients; nor has it reported the view of basic scientists who have published papers containing relevant data.

The newspaper appears to dismiss the views of such individuals on the grounds that they are part of what it describes as the "HIV industry". **Maxine Clarke**

Fusion experiments give lift to future funding prospects

Washington. Physicists at the University of Princeton are this week starting to analyse data from a series of record-breaking experiments which they hope will help to establish the commercial potential of nuclear fusion.

The experiments were carried out at the Princeton University Plasma Physics Laboratory, which last week sustained a power output from fusion of as much as 6 MW for three-quarters of a second. They were described by O'Leary, the Secretary of Energy, as "the most significant achievement in fusion energy in the past two decades."

Previously, the record for fusion power output was 1.7 MW, set at the Joint European Torus (JET) facility at Culham in Britain in 1991. But where the JET experiments used a plasma of 90 per cent deuterium and 10 per cent tritium, Princeton used a much more powerful 50:50 mixture of the two hydrogen isotopes, similar to that likely to be needed in commercial fusion reactors.

Ron Davidson, director of the Princeton laboratory, says that the duration of the fusion event is restricted by the ability of his equipment to store and deliver the 25 MW of electricity needed to power the copper magnets. These suspend the plasma in space as it is heated to temperatures of more than 100 million degrees Kelvin.

The 1 GW International Thermonuclear Experimental Reactor (ITER), being planned for an unspecified site by the United States, Japan, Europe and Russia as the next fusion prototype, will use superconducting magnets which require far less power. As a result, it is hoped that ITER will be able to sustain a continuous and energy-efficient fusion reaction.

A programme of around a thousand fusion "shots" at Princeton will continue through this winter, providing vital data on the control of the plasma, on ways of removing heat from it, and on the best choice of reactor materials. The experiment will also study the effects of alpha radiation from fusion on the behaviour of the plasma itself.

The Department of Energy, which provides the Princeton facility's $80 million budget, hopes that publicity from the experiments will help secure fusion funding in Congress, where sceptics complain that a fuel source which will yield no commercial energy until at least 2040 gets more research support than any other.

Congressional aides said that the fusion budget is unlikely to come under any real pressure next year. But laws have already been passed which make future United States support conditional on the ITER partners agreeing on an international framework for fusion research. **Colin Macilwain**

NSF opens network to Russian scientists

Washington. **Another barrier left over from the Cold War fell last week when NSFNET, the scientific information network of the US National Science Foundation, began carrying data from the former Soviet Union (FSU) for the first time.**

The decision to allow such traffic was taken by the National Science Board, and will considerably ease electronic mail and data exchanges between FSU scientists and their US counterparts.

From its inception in 1986, NSFNET, which connects approximately 10 million users in federally sponsored research laboratories, colleges and universities, and forms one of the main backbones of the global network Internet, has had a policy of refusing traffic from FSU states.

As a result, although Soviet messages could find alternative routes through Internet, Steven Goldstein of the NSF's networking and communications research division estimates that at least 30 per cent of US sites were not directly accessible to FSU scientists. These scientists should now should be able to reach all those sites by electronic mail directly.

At the same time, the National Aeronautics and Space Administration (NASA) has unveiled plans to link a group of Russian space scientists directly to its own NASA Science Internet — another major backbone of Internet — by early January. A satellite link will connect the Ames Research Center near San Francisco to Russia's Space Research Institute (IKI), which will act as the hub of a Russian network linking nine additional space research facilities.

The system will support collaborative projects in areas ranging from astrophysics to space life sciences. But its use will be restricted to scientists working on those projects. The US Department of Energy (DOE) plans a similar connection between its Energy Sciences Network and several Russian sites later in 1994.

Both the NASA and DOE efforts should benefit from work being completed in Moscow on wiring the city's leading research centres with fibre optic cable for high-quality communications. Funded by philanthropist George Soros's International Science Foundation (ISF), the work is expected to be completed next spring.

Tony Reichhardt

Lobbyists prompt US breast cancer campaign

Washington. Inspired by the success of the grass-roots AIDS lobby, interest groups representing breast-cancer sufferers have persuaded the Clinton administration to launch a national strategy to combat the disease, ranging from fundamental research to enhanced screening programmes.

Breast cancer strikes one out of every nine American women. More than 180,000 cases are diagnosed in the United States and about 46,000 people will die from the disease each year.

About two hundred experts and leaders representing all interests within the breast-cancer community met at the US National Institutes of Health (NIH) this week to thrash out the details of a national plan of action.

The NIH conference brought together high-ranking public officials from the Clinton administration, including the Health Secretary Donna Shalala and the Surgeon General Joycelyn Elders, as well as members of congress, senior NIH officials, researchers and representatives of patient advocacy and support groups.

Ruth L. Kirschstein, deputy director of NIH, says that the aim of the conference was to bring together people with "many diverse points of view, needs and interests". Rather than representing any one interest group, she says participants were expected to speak for "the population of the United States, as this is a concern that men as well as women should have".

Participants, divided into smaller groups, were asked to identify issues and goals of national priority in six broad areas, ranging from consumer education to "the translation of research into practice".

President Bill Clinton promised to convene a special conference to develop a national strategy to fight the disease after a meeting with leaders of the National Breast Cancer Coalition (NBCC) during their march on Washington DC in October.

Organizations such as the NBCC are taking their lead from the AIDS lobby, which has become a powerful political force built largely on grass-roots efforts. The strategy seems to paying off: the National Cancer Institute will spend an estimated US$263 million on breast cancer research in the fiscal year 1994 (which started on 1 October), an increase of 34 per cent over 1993.

October was also the month when the President's Cancer Panel special commission on breast cancer delivered a 26-page report in which it said that high priority should be given to research on the causes of breast cancer and prevention, to newer imaging techniques and to the continued development of more specific, less invasive and less toxic methods of treatment.

The commission also emphasized the need to promote the use of screening mammography and clinical breast examinations among older women, women from rural areas, women of lower socioeconomic status and women of certain racial and ethnic populations who, typically, are under-represented. To achieve these goals, the commission recommended that the federal government should spend "no less than US$500 million" annually on breast cancer research.

Nancy G. Brinker, chairman of the special commission, and founding chairman of the Susan G. Komen Breast Cancer Foundation in Dallas, Texas, says she hopes the commission's report will serve as the centrepiece for the strategic plan. "To re-invent this process would be a serious waste of money and effort", she says. **Diane Gershon**

Rival Boston hospitals set to merge

Washington. Two of the United States' most medically sophisticated and fiercely competitive hospitals have announced that they are to merge. The merger is designed to meet the challenge of health care reform that is being driven by an unrelenting determination to cut the cost of medical care (see *Nature* **366**, 200–202; 1993).

The Massachusetts General Hospital (MGH) and the Brigham and Women's Hospital, two Harvard-affiliated teaching hospitals located just blocks from each other in the centre of Boston, will attempt to consolidate resources in ways that have yet to be determined.

Their goal is to be able to compete for patients and research physicians against lower-cost community hospital consortia that do not specialize in teaching, research and the highest of high-technology medicine. It is a sensible alternative to allowing the two hospitals to continue an intense rivalry that dates back to the 1800s.

The merger is seen as a precedent for other leading US research hospitals that will have to form some sort of alliances in order to survive health care reform, regardless of which of the many legislative options eventually is enacted into law.

The MGH has already formed a new governing entity, the Massachusetts General Hospital Corporation, which will include the McLean psychiatric hospital and the Spaulding hospital for rehabilitation.

The corporation will be headed by Samuel O. Thier, currently president of Brandeis University and past-president of the Institute of Medicine of the US National Academy of Sciences. Thier received much of his own medical training at the MGH.

One likely result of the merger will be a gradual reduction in the number of staff at both hospitals, consistent with the current trend in 'downsizing' hospitals across the country. The challenge of course, is to consolidate resources without sacrificing the quality of research and care provided by these and other university research hospitals across the United States.

Barbara J. Culliton

High bids for second-hand space suits

Last weekend, as the Russian people voted 'yes' to President Boris Yeltsin, his free market democracy was in evidence at Sotheby's in New York when more than 200 artefacts from the Soviet space programme went under the auctioneer's hammer.

Alexei Leonov, the first man to walk in space, watched his training space suit (see right) fetch $255,500. But it was not only hard cash that motivated the sale. "It permits an exchange of cultural values," said Leonov.

Western museums and historians saw the sale as a chance to set the record straight. "If you look at museums of science and technology around the world, there's a lot of NASA material. There's virtually nothing from the Russian space programme," said David Redden, a senior vice-president of Sotheby's and the organizer of the sale.

Nonetheless, profits from the sale will certainly be welcomed by those surviving on state pensions devalued by inflation.

Seth Joel/Science Photo Library

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

The widow of Yuri Gagarin, the first man in space, made one space enthusiast very happy when she put the orange space suit he used to train for his historic flight (above, background) up for sale.

Diaries detailing the changing fortunes of the Soviet space programme and even the title deeds for two 'buggies' still on the Moon — to be paid for on delivery — all added to the total of $6.82 million fetched by the collection. □

Physicists want 1996 link-up with LHC

Washington. High-energy physicists in the United States are working on a new strategy for their field which could lay the groundwork for a substantial US involvement from 1996 in the Large Hadron Collider (LHC) being planned by the European Laboratory for Particle Physics (CERN) in Geneva.

"It is perfectly possible for us to provide an intellectual contribution [to the LHC] now, with the expectation of money to follow later," says George Trilling of the Lawrence Berkeley Laboratory in California, one of the many physicists having to come to terms with the recent death of the Superconducting Super Collider (SSC, see *Nature* **365**, 773; 1993),

At present, the several hundred US scientists who work at CERN on salaries paid by US organizations do not have to pay for access to its experimental facilities. This arrangement has been based on expectations that a reciprocal arrangement would have given Europeans access to the SSC.

With the demise of the SSC, a new arrangement is being sought. The 19 European governments who make up CERN would like that arrangement to extend to US participation in the construction and operation of the LHC.

One likely area for US involvement is in the design and construction of the LHC's huge detectors. Gil Gilchriese of Lawrence Berkeley says he does not think the United States will build its own detector. Instead, he thinks it is likely to make "a significant investment, of up to 20 per cent" in the two existing European detector collaborations from 1996. (That is the earliest date at which the cumbersome US budget process can be expected to yield new money for any international collaboration.)

In the meantime, US high-energy physicists are fighting to save some of the knowledge and expertise built up at the SSC Laboratory in Texas.

Scientists involved in the detector work want $15–$20 million in the current financial year to complete their research. But John Peoples, director of Fermilab and now also director of the SSC, says support at that level is "out of the question".

"Congress doesn't see it as part of the orderly termination" of the SSC, says Peoples. Some support may be forthcoming, he says, but only if Congress is convinced that progress is being made in saving money elsewhere as the project is shut down.

Hazel O'Leary, the US Secretary of Energy, last month asked her High Energy Physics Advisory Panel (HEPAP), whose main advice up to then had been "build the SSC", to set up a subpanel to report next May on a future strategy. O'Leary, in turn, is due to report to Congress in July.

After a preliminary discussion, Stan Wojcicki, the head of HEPAP, has told O'Leary that three issues need immediate attention. The first is that some of the $640 million allocated in the current financial year for the orderly closure of the SSC should go to support detector research and development, so that research already completed can be documented properly.

Second, HEPAP would like between $20 and $30 million of the budget for fiscal year 1995 (which begins next October), now under discussion by the administration, for detector research. This would put US physicists in a good position to participate in the LHC at a later date. And the panel also wants an end to the recent decline in the non-SSC high-energy physics budget.

Once these issues have been addressed, HEPAP says it will look at the prospects for longer term involvement with the LHC. O'Leary has told the panel that she wants it to focus on the options for creating "a truly international framework" for future work in high-energy physics.

In an effort to involve more physicists in the discussion, the particle and fields division of the American Physical Society is planning its own effort to redefine priorities. Subpanels will be set up, says Roberto Peccei, the head of the division, to produce a preliminary assessment for HEPAP in May, and a final report the following January. **Colin MacIlwain**

CERN views a suitor with caution

Geneva. As the United States prepares plans to salvage what it can from the wreckage of the Superconducting Super Collider (SSC), the European Laboratory for Particle Physics (CERN) in Geneva is considering how its own equivalent project might benefit.

Christopher Llewellyn-Smith, who takes over from Carlo Rubbia as CERN's director-general next month, says it is too soon to start making specific plans for bringing US particle physicists on board the proposed Large Hadron Collider (LHC). Formal discussion are unlikely to start for at least six months.

But talks are already taking place between European and US scientists on areas of potential collaboration. When the dust eventually settles, the most likely outcome is that the United States will be brought into the LHC programme through some form of special cooperative agreement.

One problem hampering negotiations is that the LHC is not likely to receive approval from CERN's 19 member states until next summer at the earliest, and possibly even later. Another is that the US Department of Energy's High Energy Physics Advisory Panel, which is currently drawing up plans for post-SSC particle physics in the United States, is not due to report until May (see above).

The CERN council meets in Geneva this week (16 December) to discuss the final LHC budget proposals, which have already been cut back in response to criticism from member states. Germany in particular has expressed concern that the cost of the LHC, which now stands just above SFr2 billion (US$1.34 billion), has increased around 10 per cent in real terms from initial estimates in 1989.

Llewellyn-Smith, backed by most member states, wants the proposed modified budget to be approved independently of any potential agreement with the United States, even though such an agreement could bring in extra money. At the same time, he is keen to find ways of making use of the scientific and technological expertise already built up on the SSC project. "Specific collaborations negotiated independently with the Americans could allow us to expand on the experiments we can plan," he says.

Luciano Maiani, director of Italy's National Institute of Nuclear Physics and the country's delegate to the CERN council, agrees. The LHC should be open to international collaboration in every possible way, he says. "But we can't afford to make approval of the LHC conditional on obtaining outside money."

One possible arrangement is for the United States to become an 'associate member'. But this would require a rule change, because such a category does not yet exist.

Another option is for the United States to becomes a full member of CERN. But this is unlikely to be acceptable to current member states. One reason is that the annual membership fee, which is based loosely on a country's gross national product and would reach the limit of 25 per cent of its budget allowed by CERN as a contribution from a single country; this might be more than the United States would be willing to pay.

And even if it did feel able to afford this sum, there is likely to be further opposition. As a full member, the United States' budget contribution would be the largest of any member state, an uncomfortable prospect in the eyes of some countries, such as Italy, that wish to preserve CERN's status as a European organization. **Alison Abbott**

Congressmen in last-ditch THORP appeal

Washington. A number of leading members of the US Congress have written to John Major, the British prime minister, asking for any decision on starting up the thermal oxide reprocessing plant (THORP) in Cumbria to be delayed until a new public inquiry is held into the need for the multi-billion pound plant.

The letter has also been sent to US President Bill Clinton and the Japanese Prime Minister Morihiro Hosokawa, arguing that the operation of the plant would "contravene the stated policies of Japan and the United States" by adding to the world's plutonium stockpiles.

Signed by senators Jeff Bingaman (Democrat, New Mexico), Dale Bumpers (Democrat, Arkansas) and Tom Harkin (Democrat, Iowa), as well as 30 members of the House of Representatives, the letter calls on the Clinton administration to help Britain and Japan renegotiate existing THORP agreements.

A decision on whether to start up the nuclear reprocessing plant is expected soon, although continuing delays are fuelling speculation that the British Treasury — whose concerns about uncontrollable costs are reported to have been the main reason for delay so far — is fighting a late rearguard action against approval.

A group of congressmen led by Pete Stark (Democrat, California) and Joseph Kennedy (Democrat, Massachusetts) have also written to Hazel O'Leary, the Secretary of Energy, asking her to turn down a request from Switzerland to send fuel to THORP for reprocessing.

The United States has full control over the Swiss material under supply agreements between the two countries, as it has over Japanese fuel. Japan, however, has negotiated a waiver removing the requirement that everything it does with its fuel requires that explicit approval is given by the US Department of Energy.

Despite pressure from environmentalists and others, Clinton has declined to take a public position on THORP, arguing that US intervention might offend both Britain and the plant's largest customer, Japan.

But opponents of the plant are suggesting that one factor behind Clinton's reluctance to criticize THORP is a feeling among some US officials that Japan will eventually need its own nuclear weapons capability, and that this would also be in the best interests of the United States. Plutonium from THORP would be needed to achieve this goal.

Paul Leventhal, for example, president of the Washington-based Nuclear Control Institute, says that the plutonium surplus that will be acquired by Japan if THORP comes on stream "makes no sense, except to open up a weapons option". He claims that the development of nuclear weapons by Japan "is not necessarily regarded as a danger".

But Japanese officials vigorously deny this claim, and others support their arguments that the allegation is far-fetched, given overwhelming public antipathy towards nuclear weapons in Japan. "I would never accuse the Japanese of having any intention of making nuclear weapons," says Andrew Duncan of the International Institute for Strategic Studies in London.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

THORP: still waiting for a government decision.

Meanwhile the Danish Environment Minister, Svend Auken, has urged Britain not to go ahead with the operation of THORP because of the 900 per cent increase in radioactivity that would be emitted into the sea as a result.

A spokesman for Britain's Department of the Environment rejected charges that the increased emissions would pose a serious risk to the environment, arguing that THORP's emissions would represent "a 900 per cent increase from practically nothing".

But speaking at a meeting in Copenhagen last week attended by representatives of the eight European countries bordering on the North Sea, Auken said that there was a "strong wish in the North Sea countries that the decision will be a no".

Colin Macilwain

UCSF goes solo with venture fund

San Francisco. The University of California, San Francisco (UCSF), is planning to establish its own venture capital fund to help develop faculty inventions into commercial products, following the apparent collapse of plans for a state-wide university system.

The central administration of the University of California had proposed setting up a commercial company to develop inventions from all of its nine campuses. But many faculty members objected to the proposal, claiming that it would reduce the authority of individual campuses, dampen support for university research and drain patent income away from education.

A report on the proposal, due to be published next February, is now expected to advise against the system-wide plan. But it will suggest that individual campuses expand their technology transfer offices independently, providing both a source of income and a means of transferring the results of research to society.

Joseph Martin, the chancellor of UCSF, is already planning to expand the role of its licensing office. For example, this would take on responsibility for compiling a list of promising inventions on the campus, and for building relationships with both faculty members and potential commercial collaborators.

Martin says he also is considering a UCSF-managed venture capital fund to nurture inventions not quite ready to attract commercial support. The technology would be developed at the university, probably inside special 'incubator' laboratories. Capital is likely to cpme from both US and Pacific Rim companies, as well as government agencies and venture capitalists, he says.

Martin acknowledges the need to deal with faculty concerns. In particular, he says, the new arrangements must not take faculty members away from teaching and research, or result in their work being driven purely by commercial considerations. Also, he says federal funded research should be separated from commercially-orientated work.

But not all faculty members are happy, some fear that venture capital funding will harm its central mission of unfettered research and teaching. Keith Yamamoto, vice-chairman of the faculty of biochemistry and biophysics, says he supports UCSF's efforts to expand its patenting abilities and build alliances that would bring commercial resources back to campus.

But Yamamoto says UCSF should stop short of starting companies. "Technology transfer is critical and valuable," he says. "But it is dangerous for universities to become involved in funding or encouraging faculty to set up companies based on their inventions."

Martin says UCSF will study the implications of stock ownership, and intends to form a faculty advisory committee. But he argues that, under certain guidelines, the institution should be able to share the ownership of ventures with both faculty members and outside investors.

Martin has asked an advisory committee of prominent industrialists to investigate technology transfer mechanisms, looking at the experience of other universities.

Sally Lehrman

Pilot scheme gets go ahead for southern Europe

Heidelberg. The council of the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany, last week approved in principle a radical new scheme to set up regional research groups in the countries of southern Europe.

But a decision to increase the EMBL budget sufficiently to accommodate this scheme, as well as supporting the growth of outstations in Cambridge (England) and Grenoble (France), could not be reached. Another meeting has now been arranged for mid-February, when it is hoped that a compromise will be agreed under which the budget will grow by about two per cent in real terms.

Kafatos: new DG.

Despite the financial uncertainty, Fotis Kafatos, EMBL's new director-general, is planning to press ahead with the launch of a pilot scheme for regional groups. Four groups in Italy, and one in Spain, will be established at selected host institutes during next year.

Each regional group will be led by a young investigator who will preferably be a non-national of the host country, and will also include a technician, a postdoctoral student and a postgraduate student. Such a group will cost EMBL around DM475,000 (US$150,000) a year for salaries and running costs.

Group leaders will be appointed for five years, and host institutes will be expected to consider that person for a post in their own institute within their fourth year. As such, the scheme offers the only opportunity for a young molecular biologist to work independently across European borders.

Selection of host institutes will begin early in the new year. Sir John Kendrew, EMBL's first director-general, will head the selection team, which will also include heads of the councils and finance committees of EMBL and the European Molecular Biology Organization, also based in Heidelberg.

If this pilot scheme — which will be paid from EMBL's own budget, increased or not — is successful, Kafatos wants to expand it very widely, and very quickly. Provisional plans are to add three additional groups in Italy in 1996 and three more in 1998 "to consolidate the catalytic effects begun in the pilot phase". There are also plans to build up groups in other countries, particularly those in the south of Europe.

Future development, however, depends not only on the proven success of the pilot scheme but also on acquiring funds from external sources. Fund-raising negotiations will begin next year.

The new scheme reflects Kafatos's commitment to improving conditions for science in the southern European states. But he says also that the philosophy of regional groups in different countries was always at the heart of the original EMBL mission, defined in the early 1970s, to promote molecular biology in member states as well as running a central facility.

The principle of the scheme has the wholehearted support of most staff at the laboratory. But this support is coupled with concerns that it may compete with resources in the main laboratory, which employs around 750 scientists. After years of plenty, central funding has become squeezed in the 1990s. For example, equipment money was cut back by 20 per cent this year.

Some claim that the scheme has political roots. Last year Italy threatened to withdraw from EMBL, claiming that too few Italians held staff scientist positions and that it was therefore not getting 'fair returns'. Spain added its own voice of complaint in September, saying that it was also a victim of this perceived inequality (see *Nature* **365**, 201; 1993).

Despite denials from Kafatos that the choice of Italy as primary site for the pilot scheme was the result of political pressure, this perception remains in the air both inside and outside the laboratory. And Italy has certainly been sufficiently placated to withdraw its threat to pull out.

Kafatos knows that he has taken over EMBL at a difficult time. In a climate of ever-increasing financial constraint in Europe, the central laboratory in Heidelberg has been growing steadily — Kafatos has now made a decision not to push for further growth — and commitments have been made to expanding two outstations.

These are the European Bioinformatics Institute (EBI) in Cambridge, which has managed to acquire most of its support from external funds (see *Nature* **361**, 198; 1993), and the Grenoble outstation, which uses local synchrotron facilities for determining molecular structures.

The European Synchrotron Radiation facility (ESRF) will start operating routinely next year, and the EMBL council believe it must make full use of this new resource by expanding the inadequate laboratories that are already there, and doubling the number of staff from 30 to 60. France last week committed itself to providing a building, but equipment and running costs are still to be found.

Alison Abbott

Russian foundations under pressure to cooperate

Moscow. Both financial and logistical pressures are increasing on the various independent foundations established over the past two years to support Russian science. As a result, the long-held dream of the Russian Ministry of Science and Technical Policy of coordinating the activities of these foundations may be coming true.

The largest and most effective body is the International Science Foundation (ISF), established last year by the Hungarian-born financier George Soros. But the foundation is currently unable to meet all the requests deserving support that it has received, and neither can the Russian Foundation for Basic Research (RFBR).

As a result, Vladimir Skulachev, the chairman of the Russian advisory committee for the ISF, told a meeting in Moscow two weeks ago that the foundations have reached an agreement — at present unofficial — to exchange lists of applications that they would like to fund.

If a particular project is present on both lists, said Skulachev, negotiations might take place about how it could be jointly financed. Skulachev estimates that coordinating activities in this way could save the RFBR about one-third of its funds, and the ISF about $8 million; such savings would allow significantly more projects to be financed.

The ISF is hoping to reach a similar agreement with INTAL, the aid body being set up through the European Commission in Brussels.

Alexander Goldfarb, director of the ISF's New York office, says that coordinating the activities of the various foundations would not only increase the effectiveness of the way research funds are spent but would also ensure that applications are made to the most appropriate body.

"In many cases we receive applications for projects which deserve a grant, but which we cannot approve since we do not officially finance its topic," says Goldfarb. "We would like to have an agreement with the other foundations that would enable us to forward applications to where they might be approved."

Meanwhile, the shortage of funds at the ISF is becoming increasingly worrying. Although discussions continue on a possible contribution from the US Congress, the search has so far been unsuccessful.

There have, however, been some indications from Boris Saltykov, the Russian minister of science, that the foundation could receive support from his ministry. Skulachev says that it will be "very disappointing" if the foundation has to close down because of the lack of funds.

Vladimir Pokrovsky

Hong Kong scientists forge links with China

Hong Kong. While Britain and China continue to clash over the future of Hong Kong after the British colony is handed over to Chinese rule in 1997, scientists in Hong Kong's three universities are already engaged in building links with mainland China.

Next month, the Hong Kong University of Science and Technology (HKUST), established only two years ago, will open an institute to carry out research on Heinan, an island off the coast of southern China with a population of nearly 7 million.

Two other institutions — China University of Hong Kong, in the northern part of the colony, and the University of Hong Kong, on Hong Kong island — are also involved in establishing links with the mainland. This process has recently been boosted by the return of a large number of Chinese academics after many years overseas, particularly from the United States.

One of these is HKUST's vice-chancellor, Chia-Wei Woo, who was born in Shanghai, and spent his childhood in Hong Kong and Taiwan. He went to the United States at 17, and eventually became president of San Francisco State University.

Woo says he is not worried about the Chinese takeover in 1997. "The mentality that Hong Kong is on the fringe of the British Empire has to change," he says. "We are the most important city for 600 million people south of the Yangtze River."

The Heinan Institute will be staffed by 30 faculty members from the university, and will focus on six areas of research: agriculture and fisheries; behavioural science; economics and finance; environment; infrastructure; and technology transfer. The institute will also teach 180 local Heinan government officials how the Hong Kong government is run and how businesses operate in the colony, in order to "liberalize their thoughts" says Woo.

HKUST: a modern campus establishing links with Heinan Island.

Heinan is an island of beautiful beaches and verdant scenery, and particular emphasis will be placed on environmental research. The university will set up an air-quality monitoring system on the island and a ground satellite receiving station at HKUST. Scientists will be able to observe the island and surrounding sea using US and Japanese Earth-monitoring satellites.

The Chinese University of Hong Kong was formed 30 years ago by the merger of three colleges staffed by Chinese academics who fled the communist takeover in 1949. The university has been developing links with mainland China since a visit by a delegation from the Chinese Academy of Sciences in 1978.

The university now has links with a hundred universities and 70 research institutes in China, each described as "pockets of excellence" by Kenneth Young, dean of the graduate school. Many US and European officials seek its help in establishing contact with academics in China, he says.

Even the University of Hong Kong, formed more than a hundred years ago and steeped in British colonial tradition, is expanding its links with the mainland. "For the average person in the street there is no border at the Northern Territories," says T. J. Biscoe, the university's deputy vice-chancellor. "The real border is socio-economic, and extends as far as the Pearl River basin." Senior Chinese academics from the mainland regularly visit the university to establish new areas of cooperation, and Biscoe says it is "inevitable" that links will expand.

Hong Kong's universities have been undergoing rapid expansion since 1990, under a programme aimed at strengthening graduate research. Many Chinese scientists have been recruited from leading institutions in the United States and Europe.

At HKUST, for example, most of the 360 faculty members are of Chinese origin. Many, like Woo, were born in mainland China or Taiwan, and later moved to the United States.

These academics still maintain strong links with the United States. Some still hold positions in US universities, and all are connected to the INTERNET computer network, which keeps them in touch with the latest developments in the United States and elsewhere. But at the same time many have a desire to establish research links with what they describe as their "homeland".

The same trends can be seen at the other two universities, although to a lesser extent. Wang Gungwu, vice-chancellor of Hong Kong University, says that "judging from the faces at the welcoming parties", slightly more than half of the new recruits to the faculty are Chinese, largely from the mainland and Hong Kong. This marks a new trend for the university, which traditionally recruited many of its faculty members from Britain and elsewhere in the West.

David Swinbanks

UK urged to back electronic publishers

London. Electronic publishing in the United Kingdom gained further support this week from a report by the Joint Funding Councils' Libraries Review Group. The report recommends that £20 million be invested over the next three years in pilot schemes designed to stimulate the creation of an electronic library service.

The review group was chaired by Sir Brian Follet, vice-chancellor of the University of Warwick. It suggests that £2 million be spent by the four councils — the Higher Education Funding Council for England, the Scottish Higher Education Funding Council, the Higher Education Funding Council for Wales and the Department of Education for Northern Ireland — on developing refereed electronic journals..

The problem with electronic journals is not their feasibility, says the report, but their status in the scientific community. Publishers are holding back because electronic journals are not widely perceived as the best place to find or publish top quality papers.

To encourage scientists to publish more electronically, the report urges the funding councils to accept refereed electronic articles on the same basis as printed ones in their next research assessment exercises.

A further £3 million should be used to look into electronic 'bespoke' publishing, the report says. The idea is that an individual user should have access to all the information that goes into a traditional book; but instead of having the whole book, the user would be able to download individual chapters as and when required.

So far academic publishers have shown little interest in this scheme, possibly because of copyright problems. But the report recommends access controls should form an integral part of any framework established for such 'on-demand' publishing.

Other elements of the £20 million package include a study of the costs to institutions of the facilities offered by JANET and its successors and a feasibility study of the British Academy's proposed Arts and Humanities Datacentre.

Information technology, says the report, has the potential to revolutionize libraries. Its recommendations are to assess how technology can be channelled into the 'virtual library' of the future.

Fiona Gammie

On changing water into wine

SIR — In advocating the memetic basis of religion, M. Vaneechoutte[1] has overlooked or deliberately ignored the written record upon which Christianity is based. Jesus Christ was a real man who claimed to be the son of God and whose teachings were heard and whose words and actions were recorded for posterity. Neither Jesus nor his contentious teachings were conjured up by the human intellect in order to endow an individual with certainty about his or her fate (excepting perhaps the fate of persecution suffered by the early Christians). On the contrary, Jesus taught that no one has ever seen God[2], and he left mankind with no reassurances about its fate[3]. Hence Christianity as a religion is far from being memetically based.

Part of the basis of my religious belief stems from a physicochemical change that occurred nearly 2,000 years ago and that, to my knowledge, has not been repeated since. A wedding party at which Jesus was present ran out of wine, and Jesus's mother urged him to do something about it. Somewhat reluctantly, he told the servants to fill some stone water jars with water. Upon tasting it, the bridegroom was surprised that such good wine had been held back to the last[4]. I imagine that for people of that nonscientific age, the changing of water into wine must have been an impressive feat of 'magic', which must be why the event was recorded. Changing water into wine may seem insignificant in comparison to the subsequent deeds ascribed to Jesus in the New Testament, but all the physicists and chemists in today's world together could not begin to generate ethanol, a carbon-based compound, from hydrogen and oxygen atoms. When it is considered that the 'religion' of science holds that principles of physics and chemistry govern biological systems, it follows that anyone capable of transforming 'inanimate' matter would probably also be able to walk on water or repair matter within the human body, as Jesus is said to have done.

Being a follower of Christ comes down to a question of faith, not memetics. Just as I have faith in science as being the pursuit of truth, thanks to the written record in the Bible I also have faith in Jesus, a person who taught that the pursuit of truth is worthwhile *and* that we are God's creation, not He ours.

R. A. Savidge
University of New Brunswick, Fredericton, New Brunswick, E3B 6C2, Canada

SIR — Vaneechoutte[1] postulates a memetic basis for religion. Unfortunately his hypothesis is based on a false premise. He asserts that "the essence of religion remains the influencing of the future: the major reason for behaving according to the rules is that one can propitiate the deity by doing so". This is a clear confusion between the objectives of science and religion.

When I stand in a pulpit on a Sunday, I preach the major Methodist doctrine that we are saved by grace through faith alone. Simply, this doctrine means that no action(s) of our own — whether good or bad — can influence God's actions towards us. God chooses to bless us because it is his nature to do so. However, when I work as a research material scientist throughout the week I obey the rules because I can propitiate my employer and my scientific peers by doing so.

Vaneechoutte argues that religion is an emergent characteristic which is a meme which reduces fear, uncertainty and depression. It may be an emergent characteristic but its challenges often increase fear, uncertainty and depression. Indeed, the practice of Christianity is almost certain to increase the need to cope with these effects.

For example, the fundamental principle of the Christian religion is the commandment of Christ that his followers should love their neighbours as themselves. Simply, this means sharing the pain, risks and problems of neighbours. When pressed to define a neighbour, Christ explained that they are those one most dislikes. He then illustrated this by the unnecessary, expensive and risky, compassionate care of a Samaritan for a Jew (that is, religious and racial opponents) and called for extra service to Romans (hated representatives of the occupying power).

Vaneechoutte recognizes that his hypothesis cannot be equated with Buddhism, so he asserts that Buddhism is "generally not considered a religion". It is true that it is not necessary to believe in a deity to practise Buddhism, so some people have attempted to categorize it as a philosophy and not a religion. However, most of its followers and most others consider Buddhism to be a religion that may be practised by deists and atheists. Indeed, many Buddhists are monks whose only livelihood is the support of other Buddhists. Vaneechoutte has confused theism and religion in his attempt to construct a mechanistic explanation of the human desire for religion.

Perhaps *Nature* should obtain peer-review of theologians before it again publishes a piece about religion?

Richard S. Courtney
31 Rivelands Road, Swindon Village, Cheltenham, Glos GL51 9RF, UK

1. Vaneechoutte, M. *Nature* **365**, 290 (1993).
2. Bible (John 6:46).
3. Bible (Mark 13:19; John 5:28–39).
4. Bible (John 2:1–11).

Tools, not laws

SIR — Andrew Ehrenberg's article "Even the social sciences have laws" (*Nature* **365**, 385; 1993) may confirm the prejudices of many natural scientists against the social sciences. Ehrenberg says that major brands have more loyal customers than minor brands, and explains this by reference to a "law" called double jeopardy (DJ): those who only know the major brand buy it every time, as they do not know about equally good alternatives. That makes them more loyal than the customers of minor brands, who split their purchases between the two brands.

If DJ is a law, then it is also a law that twice two makes four. A parallel, in classical ("if p, then q") terms would be "if a bachelor, then unmarried". DJ is nothing but a logical implication of the assumption stated. A is a well-known brand, B is less well known but "equally good". It must be assumed that all who know B also know A.

It is implied in this information that A's customers will be more loyal than B's. Those who buy A do not know about B, whereas those who sometimes buy B by definition buy A as often. The only way in which we can confirm that the two brands are "equally good" is by letting a sample of the customers familiar with both make the choice in the supermarket.

If we consider DJ a falsifiable hypothesis, embodying a statement about the real world, and go out there to test it, it can only be confirmed. If we find a bachelor who is married, it does not falsify the "law" stated above; it is simply that he is not a bachelor. Likewise, if we find cases that seem to falsify DJ, the simplest explanation would be that the two brands are not, after all, "equally good", or that A-customers have other preferences than B-customers, which they were not supposed to have.

I do not know if there are social "laws". Those proposed have proved valid only if supplemented by an infinite number of conditioned clauses such as: "Valid only if not known by the population under investigation". Tireless attempts to provide nomothetic generalizations have given us much idiographic knowledge, but not the kind of reliable, stable and universally valid information about society that Boyle-Mariotte's law gives us about gases. DJ is not a law, but a tool that Ehrenberg and his colleagues can apply to the task of mapping empirical buying behaviour.

Ottar Brox
Norwegian Institute of Urban and Regional Research, Oslo, Norway

Letters submitted for Correspondence should be typed, double-spaced, on one side of the paper only.

Demise of the Texas supercollider

David Ritson

Two months ago, the superconducting supercollider was killed by the US Congress. Here, a particle physicist provides a post mortem and asks whether Europe should build its own collider.

WITH the cancellation of the Superconducting Super Collider (SSC), the US public must now try to understand how it has spent more than $2 billion effectively to create holes in the ground — and will now have to spend a further $640 million to refill them. There are an abundance of theories to account for these events.

On one extreme, many members of Congress and prestigious scientists, largely from disciplines outside high-energy physics, believed that the SSC was a vast pork barrel: inefficient, wastefully managed, with an ever-increasing price tag, an ever-receding completion date, and with the main spinoff being to satisfy the egos of a few scientists. Cancellation was essential to prevent further waste of resources desperately needed for the economy and for other scientific fields.

At the opposite extreme, other members of Congress and the executive branch of government believed that the SSC was a well-run project, within budget and on schedule. Presidents Reagan, Bush and Clinton, together with many distinguished scientists, believed both that spinoffs would be vital to the long-term growth of US technology and that termination would severely disrupt a 2,000-year-old tradition of scientific development. How was it possible, even allowing for hype and self-serving propaganda, for such completely divergent views to coexist? What follows is a brief history of the project.

In the early 1980s, a few visionary and charismatic high-energy physicists, including R. R. Wilson (then head of the Fermi National Accelerator Laboratory, or Fermilab), began to discuss the feasibility of a large proton collider that could open up a qualitatively new domain of physics, rich in new advances in fundamental understanding. Costs in the range of $3 billion were talked about. The European laboratory at CERN in Geneva had stolen a march on US laboratories with the confirmation of the unification of weak and electromagnetic forces using its proton–antiproton collider. The Europeans were starting construction of a large electron–positron collider (LEP) at CERN, and of a hadron–electron collider at DESY in Hamburg. US physicists felt that their 40-year undisputed lead in high-energy physics had been overtaken.

One project under construction in the United States at that time was ISABELLE, a small but high-intensity proton collider, at the Brookhaven National Laboratory. ISABELLE had run into technical difficulties concerning its superconducting magnets, causing considerable delay, so most high-energy physicists regarded this project as too little and too late. But George Keyworth, then scientific advisor to president Reagan, suggested to the high-energy scientific physics advisory panel (HEPAP) to the Department of Energy that should it recommend the cancellation of ISABELLE he would support its replacement by an order-of-magnitude more expensive superconducting super collider. For the Reagan administration, the motivation was largely that a superpower needed a superconducting super collider — and thus at its conception, the SSC was to be a national machine. When a large contribution from Japan was later sought, suspicion of the reliability of the United States as an international partner, Japanese funding problems and the national character of the project provided insurmountable obstacles.

World's most expensive man-made hole?

HEPAP was delighted by Keyworth's support and in 1984 the Department of Energy approved the setting up of a central design group under the supervision of URA, an association of 60 (now 80) universities that manages Fermilab. The design group was to provide research and development, design, an estimate of costs needed for site selection and a full proposal. The direction of the project was entrusted to Professor M. Tigner of Cornell, one of the most highly regarded accelerator builders in the world. Tigner assembled a group of experienced physicists and engineers from the Lawrence Berkeley laboratory and established collaborations with Brookhaven, Lawrence Berkeley and Fermilab to design and provide prototype superconducting magnets. By 1986, the group had already prepared a design report and careful cost estimates for the future accelerator. The report was widely regarded as a model of scientific clarity. By the end of 1988, after intense competition, a site for the machine near Dallas was selected by the Department of Energy, entailing the creation of an entirely new laboratory. Many subsequent problems could have been avoided if the SSC had been sited at Fermilab, which already had an accelerator complex. The ostensible reason for the choice was the superior terrain at the Dallas site, but it is widely believed that the decision was heavily influenced by the political clout of Texas.

The machine was to have a circumference of 54 miles, three times that of the LEP accelerator by this time running at CERN. In its January 1989 budget request, the Department of Energy estimated that the machine, together with experiments and laboratory complex, would cost $5.9 billion in real terms over the 8½-year construction period. Of this sum, the US government was to supply $3.9 billion, the state of Texas $1 billion and the rest was to come from other countries, mainly Japan. The $3.9 billion from the government was expected to be in large part 'new' money, but also to come from phasing out existing funding as high-energy physicists switched fields towards preparing SSC experiments. Thus in 1989 the net bill to the US taxpayer was expected to be around $3.5 billion.

In actuality, the pre-existing programme was separated from SSC funding, and the national laboratories fought strenuously and successfully to maintain and even increase their funding. The Japanese contribution, like the pot of gold at the end of the rainbow, never materialized. So even if the original budget had held up,

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Philip Bermingham

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Fermilab

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

D. Lynch-Benjamin

Participants in the battle for the SSC. From the left, Tigner, an early casualty; Schwitters, SSC's director; Edwards, Tigner's successor; Watkins,

the final cost to the US taxpayer would have been closer to $4.9 billion.

A major problem now surfaced. In the economic climate of the 1980s, the US administration believed that a programme like the SSC would be successful only if it was predominantly run by people with managerial skills obtained from the military or from industry. The role of academic physicists and engineers would be mainly to provide the designs and parameters. This view was held despite the fact that LEP had already been successfully completed and commissioned under traditional physics management on budget and on schedule. The administration classified the central design group and URA as 'academics', and thus as suspicious.

Subsequent to the site selection, therefore, bids were solicited for the prime contractor. It was believed that a direct bid from URA would not have met government specifications without "teaming" with industrial partners. Surprisingly, URA members were sympathetic to this restriction — or at least ready to compromise so as not to jeopardize the project. Edward Knapp, then president of URA, hired Douglas Pewitt, a physicist, to supervise the preparation of URA's proposal. Pewitt was a Washington insider at that time working for a scientific consulting company. He collaborated with Fermilab and central design group staff to prepare the proposal and arranged with Sverdrup, a very large conventional construction company, and with EG&G, a company that had provided personnel and management for the nuclear testing programme, to team up with URA. Pewitt further worked behind the scenes to persuade the other industrial bidders to drop out of the competition in the national interest. In this he was successful, and URA won the management contract.

The SSC director had to be chosen on a very short timescale. Unlike their European counterparts, US laboratory directors have a lifetime appointment. Thus the pool of experienced people was very small, and potential candidates were already committed to their own laboratories. Professor Roy Schwitters of Harvard University became URA's choice. Schwitters had had a distinguished scientific track-record but his previous directorial experience had been as co-spokesman for a large collaboration constructing the collider detector facility for use at Fermilab. He had also served on the programme advisory committee of the Japanese KEK accelerator laboratory. Unlike the European and US versions, the Japanese ran their laboratory with a very small permanent staff and relied heavily on industry to provide research and development, design, construction and operation. Schwitters had been impressed by this model and was ready to adopt substantial aspects of it in running the SSC. But even with the tight control that the Japanese government has over its industries, KEK required roughly twice the annual budget of an equivalent European laboratory.

Schwitters was inexperienced and, perhaps not surprisingly for somebody thrown into this maelstrom, turned out to be a 'hands-off' director in a situation that would have taxed somebody of much greater experience. He was now faced with creating a completely new laboratory on the fringe of suburban Dallas which had to proceed very fast to construction. It was agreed that Tigner should be a member of the laboratory directorate, where he could oversee the construction of the SSC. But some mix of the Department of Energy and URA did not like this idea, so Schwitters withdrew the offer and offered Tigner a lesser position. Not surprisingly, Tigner resigned. The attitude of suspicion to academics and the central design group became widely known within the group. One leading figure in the group summed up the prevalent attitude to me over lunch. "We couldn't have done everything wrong, after all the project has been approved." Only a small fraction of the SSC's important nucleus of experienced designers and engineers chose to stay with the project.

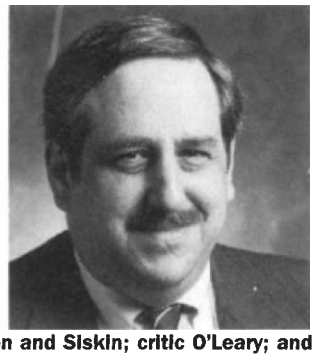
Dr Helen Edwards, who had just successfully commissioned the Tevatron collider at Fermilab, was appointed as Tigner's successor. The next step was to assemble the staff. Established and even junior high-energy accelerator physicists and engineers were reluctant to make the transition to the 'wilds' of Texas. The directors of other national laboratories tried to retain their best staff. The actual mechanics of recruitment were inept; candidates were frequently left to their own devices rather than being made to feel wanted and welcome. Thus, despite the importance of the SSC, the programme never recruited more than a fraction of the experienced senior scientists that normally would have been considered necessary for this type of project.

Sverdrup (overseeing construction) and EG&G (supplying procurement, computer services and business support services) put in 70 and 200 personnel, respectively, jump-starting the staffing process. For a brief honeymoon period, my academic colleagues were delighted. A library or a computer centre could be ordered, and a few weeks later it would be there. But the academics rapidly became disenchanted when they came to take a closer look at the actual functioning of these entities — one compared it with a CIA sting operation, where a large dummy corporation is created out of thin air.

Sverdrup's staff was later reduced to about ten. Lockheed was then added to the team and was finally to supply a staff of 70 for systems engineering. EG&G staff tripled to more than 600 people by 1993 (and was projected to peak at 789 in 1994). Thus by 1993, more than one-third of SSC's total staff of 2,000 were loyal to outside companies. The 1993 leaked audit by the Department of Energy pointed out that the placement of EG&G personnel "provides the firm with the potential for a conflict of interest when estimating costs required for its own operations".

By the standards of other laboratories, SSC staffing was very light on experienced machine personnel and top-heavy on administrative personnel. This could result only in less efficient design and execution; it is not surprising that from this cause alone there were substantial cost increases above the original estimates.

During 1989 and 1990, the accelerator division worked on design and cost estimates specific to the site. In December 1989, with strong backing from the machine advisory committee (an outside committee with international membership) it was decided to attempt a more conservative design for the main collider rings. The changes were for a magnet aperture increased from 4 to 5 cm, a lattice that incorporated stronger focusing, and



DoE head; rival candidates Reardon and Siskin; critic O'Leary; and Johnson, congressional supporter.

an injection energy doubled from 1 to 2 TeV. These changes on their own should have increased the total cost by around $700 million to approximately $6.6 billion. By summer 1990, these and other re-estimates increased the budget to completion to around $8.25 billion.

Admiral Watkins, then head of the Department of Energy, requested URA to appoint a new project manager and to strengthen substantially the management group overseeing the design and construction of the accelerator complex. URA proposed to appoint Dr Paul Reardon, a physicist with both Department of Energy and national laboratory accelerator experience. Independently, Watkins proposed Edward Siskin, a former vice president of Stone and Webster (a large construction company in the energy field). In a compromise, Siskin was appointed to the new position of general manager with Reardon, the new project manager, under him. Under Reardon in turn came Edwards as head of a accelerator design. In a further major structural change, Watkins appointed a local Department of Energy project director, Joseph Cipriano, with strong oversight powers, who would report directly both to the Department of Energy's Director of Energy Research and to Watkins, bypassing the usual Department of Energy chain of command. As a byproduct, the Department of Energy annual appraisals no longer took place, losing an important mechanism for scientific and technical oversight.

These changes were to cause serious new problems. Cipriano's office contained about 60 permanent and 40 temporary Department of Energy staff. Intense animosities developed between his office and the SSC staff, and directly between Schwitters and Cipriano. Cipriano's officials were blamed for excessive micromanagement and for forcing an exceedingly bureaucratic structure on the SSC. In turn, department officials felt that laboratory relations were worse than anything they had previously encountered, that the SSC personnel were arrogant and belligerent, and that the atmosphere was "them versus us". Certainly, the laboratory structure was exceedingly bureaucratic, but the blame must be shared between the Department of Energy and the SSC managements. In an interview with the *New York Times*, Schwitters further aggravated the situation by calling the Department of Energy oversight "the revenge of the C students". It is strange that URA, representing as it did more than 80 universities, never sat down with Watkins to discuss this or any other concerns. Whatever might have been its attitude in private, URA's public stance was to hear no evil, to see no evil and to speak no evil.

These animosities were significantly to damage the project in the summer of this year. The Bush administration was very supportive of the SSC and was content not to probe too deeply. But with the change of administration the SSC came under intense adversarial scrutiny, with its bookkeeping systems a major concern. Hazel O'Leary, head of the Department of Energy in the new Clinton administration, in testimony to Congress expressed her deep concerns about "the negative adversarial relationship" of SSC personnel and "about a pattern of behavior on the part of laboratory management which threatens the success of the project".

The second result of these changes was for the SSC management and the Department of Energy to try to allay congressional fears by deciding from then on to present the project as on schedule, on budget, and with unchanged design and scope. To comply with this edict, SSC staff became very defensive about the details of budgets and schedules, attempting to offload portions of their budget onto other groups or to invent categories of expenditure to conceal projected increases. The cost and schedule control system had always been weak or nonexistent, and its problems were severely compounded by this schizophrenic split between reality and management expectation.

A further victim of this process was that it became difficult, if not impossible, to assess potential cost savings from tradeoffs or redesigns. Therefore the usual mechanisms for maintaining a project on budget were severely weakened. Only this summer was this problem recognized and a system of "value engineering" started.

Now to return to the history of the SSC. In spring 1991, Edwards proposed modifications to the collider vacuum systems and quadrupole apertures, was overruled by Siskin, and resigned. The changes she favoured were unanimously endorsed by the machine advisory committee, and the issue was repeatedly brought up at subsequent meetings. This summer, the changes were finally approved.

For the next 9 months Edwards was effectively replaced by her husband Don Edwards, who retired at the end of 1991. Around this time Reardon resigned as project manager and was replaced by Professor John Rees, who is highly regarded throughout the accelerator community for his experience in supervising the construction and commissioning of several machines at the Stanford Linear Accelerator Center. Rees was quietly and systematically to reorganize the groups under him to make more effective use of the available talent. But at this late stage and with the political uncertainties now surrounding the SSC it was very hard to recruit experienced senior staff.

In June 1992, the House of Representatives, perturbed by the large price-tag and by the absence of any Japanese money, voted by 232 to 181 to shut down the project. But the Senate voted in favour and a stacked conference committee restored full funding. Inevitably, the people working for SSC were very worried by this close brush with disaster and by the dimming hope for Japanese involvement.

In the light of Congress's attitude, it became even more important to maintain that the project was on budget and on time. Priorities were reordered: the conventional construction of the 54-mile main collider tunnel and the contracts for the collider magnets were pushed forward at the expense of the feeder rings required for injecting protons into the main collider. This decision seemed driven more by a desire to cross a point of no return as early as possible than by scientific prudence, which would have given priority to completion of the feeder rings to provide indespensable real-world experience on controls, operations and reliability for the "green" staff.

We now come to the events of 1993 that killed the SSC. The support of the Bush administration and by Bush himself had been massive. As an example, a friend of mine was in a congressman's office when, out of the blue, Bush telephoned from Air

Force One to elicit support for the SSC. But even with Bush's personal support, the SSC had a narrow escape in 1992.

The attitude of the new Clinton administration was much less clear. The secretary of the Office of Management and Budget and vice president Al Gore had in the past been negative about the SSC. O'Leary (the new head of the Department of Energy) testified at her confirmation hearings that she felt "less than passionately about the SSC". This probably summed up the prevailing administration attitude. But Texas was still of major political importance and there was optimism in the summer when Clinton publicly supported the SSC. But how far this support extended in private is questionable.

As to Congress, the elections and budget debates made it abundantly clear that 1993 was a new ballgame, with a very high priority put on deficit reduction. Additionally, there had been a large turnover in representatives and the newly elected congressmen supported deficit-reduction measures even more strongly.

In this climate it might have appeared advisable for the SSC to have looked for economies. A typical suggestion was to construct one large $500 million detector (instead of two) for the start of operations and to provide a real second-generation detector later. But no moves to economise were made.

The Clinton administration proposed to spread costs by postponing completion of the SSC by 3 years. The Department of Energy estimated that this would increase costs by $2 billion, assuming uncharged staffing levels and with no adjustments for a decreased workload. The SSC never seriously explored how best to accomplish economies by stretching out projects. Its hard-line strategy seems to have been based on the judgment of, among others, senator Bennett Johnston of Louisiana, the powerful chairman of the Senate Energy and Natural Resources Committee, who privately advised the SSC not to modify its stance at all. He promised to ensure passage of the full budget.

On 19 June a draft report was leaked from the inspector general of the Department of Energy criticizing subcontractor expenditures. Like an earlier report in February 1993 from the General Accounting Office, this one asserted that legitimate and mandatory oversight had been deliberately impeded by SSC staff. On 24 June the House of Representatives voted by a margin of 280 to 150 to kill the SSC.

On 30 June there were hearings before congressman Dingell of the House Subcommittee on Oversight and Investigation to investigate the allegations. These hearings were widely publicized and made the SSC a symbol of wasteful expenditure. The inspector general's attacks on improper expenditures on luxury items such as flowers and meals were very photogenic. One example was the potted plants used to enliven the sterile warehouse with its cubicled offices which housed SSC staff. A second was dinners provided to visiting committees otherwise serving in a *pro bono* capacity. These funds had been provided by URA out of its "management allowance": technically these funds were restricted and thus subject to government regulation. Had URA followed the same practice as its sister organization (AUI) which runs the Brookhaven National Laboratory and negotiated for a 'fee', the money would have been unrestricted and there would not have been a problem. One can only weep that URA, many of whose 80 universities had recently been savaged by accusations of misuse of government overhead costs, did not have the foresight to anticipate this problem either through a fee or by supplying unrestricted funds through its member universities.

More serious were the attacks on the book-keeping and cost-accounting procedures of the SSC. The system was indeed inadequate. Ironically, the items singled out for attention were the magnet prototyping done at other national laboratories which were, in fact, well-executed and cost-effective, although the paperwork involved may have been lighter than the bureaucracy would have liked.

O'Leary's defence of the SSC was less than resounding. Basically she agreed with the criticisms, blamed them on previous administrations but promised that the Clinton administration would reorganize the SSC management and that she personally would set in place mechanisms to ensure that the SSC was completed on time and on budget. Her solution was to bring in a large industrial contractor to oversee construction and to use URA to supervise design and final operations. Had the SSC continued, the cure would have almost certainly been more deadly than the disease.

In this climate, the successes and technological advances of the SSC passed almost unnoticed. Privately, Senator Johnston still reassured the SSC that the history of 1992 would be repeated. He stated that he would ensure that there were favourable hearings on the SSC before his committee, that the Senate would in late September pass the full appropriation, and that the house senate conference committee convened to resolve the differences would again be stacked in favour of the SSC. Indeed, a favourable hearing was held in early August.

But the tide continued to run against the SSC. On 1 September, the Department of Energy issued a report from its review committee on the status of the SSC. The 75-member committee was a mix of department personnel and a highly respected group of outside experts with extensive accelerator and construction experience chaired by John Scango, a deputy director for programme management in the department. During August, they had spent three weeks at the SSC site. An underlying theme of the report was the difficulty that the investigators had had in obtaining solid cost and scheduling information. They further flagged that the final figure for magnet fabrication, one of the main uncertainties, would remain uncertain until General Dynamics and Westinghouse bid on the production-line costs 2 years in the future. They believed these bids were likely to exceed the current SSC estimates, and so they labelled their final figure as necessarily imprecise. If the project was to continue unchanged, their best judgment was that it was probably a year behind schedule and that the price to completion was likely to be $1.5 billion higher than the SSC 'on-budget' figure of $8.5 billion, with a net total cost to completion of $10 billion. Allowing for $1 billion from Texas, the US taxpayer would then be required to spend a net total of $9 billion. If the project completion date was set back 3 years, the 'hard-line' SSC estimate of an additional $2 billion would have increased this to $11 billion. This is to be compared with the first budget estimate, which assumed Japanese support, of a net cost in real terms of $3.9 billion.

Initially, the scheme outlined by senator Johnston unrolled exactly as he had predicted. The senate voted in favour and the conference committee reported out the bill at the full $640 million appropriation. But then the process unravelled. The angry house membership voted by an even greater majority than in June, 242 to 183, to send the bill back to conference. The house instructed its representatives in the conference committee to terminate the SSC. This action ended the SSC early in October.

Could this outcome have been avoided? When a ship that is undermanned and poorly prepared encounters a violent storm and sinks one can never say that even if it had been fully manned and well equipped it would not have sunk. All one can say is that it would have helped.

I have written this account in the belief that there will and should be other large scientific endeavours, and in the hope that at least they will avoid the obvious errors made by the SSC. Finally, I hope that my European colleagues will understand that the CERN-LHC superconducting proton collider does not face the problems of poor management, weak infrastructure and politicization that so weakened the SSC, and that they will not be deterred by the fate of the SSC, and will proceed to investigate the TeV physics that is so important to the future. □

David Ritson is emeritus professor of physics at Stanford University. For the past 5 years he worked as a consultant for the SSC accelerator division.

India's great ambition to succeed

Never can there have been such a determination to attain improvement by science and technology as there has been in India since 1947, but success is sadly still a long way off.

INDIA is not a country, but a continent. Its population will exceed 1,000 million early in the next century. But it is also a poor country, with an average annual income of roughly US$190 in 1991. Yet in almost half a century of independence, India has never doubted that it will become, by the application of science, one of the world's leaders. It is a proud country, but it also has a chip on its shoulder.

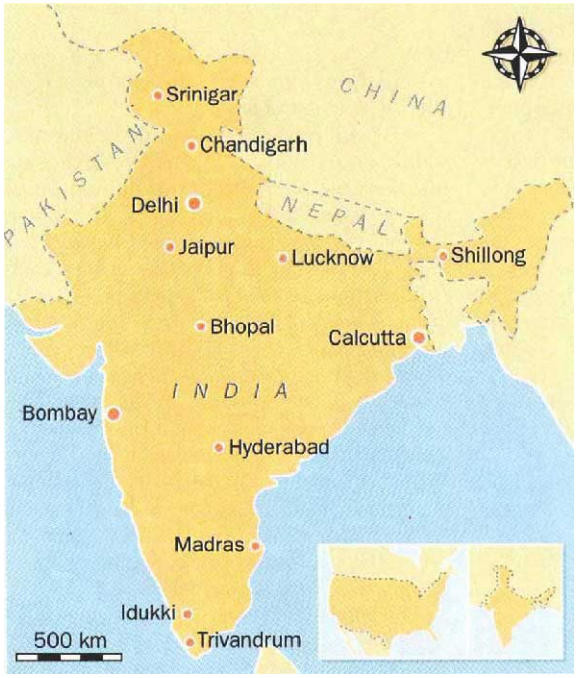
The pride is easily explained. India is the world's largest democracy, with more than 850 million people. It is also a multicultural society: there are more Muslims in India than in neighbouring Pakistan, while the Tamils of the south differ from the Bengalis of the northeast, the Sikhs of the north and the Gujeratis of the northeast not only in what they wear but in their native languages. (There is something in the jibe that, by choosing Hindi as the national language, the government of India put all people at an equal disadvantage; mere citizenship requires two languages, the ambitious must speak English as well.) People are proud that, with all the difficulties, this huge nation has held together for nearly half a century since independence in 1947.

India is also proud of its more distant past. Hardly a day passes without a reminder that the concept of (and the symbol for) the zero of arithmetic was the invention of some unnamed Hindu in the ninth century AD. Ancient texts are searched for proofs that the earliest astronomers included Hindus. The archaeology of the Indus and Ganges valleys (the former mostly in Pakistan) will further exalt pride in the past when it is more fully uncovered.

Even the palaces of the mediaeval Mogul occupation from the West, their architectural features tactlessly echoed in the monumental Lutyens buildings in New Delhi, have become symbols of past richness. The subsequent nibbling at the coastline of India by Portuguese (Goa), French (Pondicherry) and British (eventually almost everywhere) is less a source of pride than a reminder to Indians of their good fortune in having had the subcontinent to themselves since 1947. As the symbols of the British Raj fade, India becomes more like a place in which intelligent people are working out what to make of an unmarked canvas. India is an adventurous place.

One asset is the sheer scale and variety of the subcontinent (absent Pakistan, Bangladesh and Sri Lanka). From top to bottom, from the disputed border with China to the tropical waters at 8° North, India spans 28° of latitude, a quarter as much again as the north–south distance between the Canadian border and Brownsville, the most southerly town in Texas on the Gulf of Mexico. The Himalayas in the north are a great source of pride, the patches of tropical jungle in the south a marvel even to those who live there, and the land in between is rich with minerals and sufficient (provided the irrigation canals are kept working) to feed even the teeming population.

India's other outstanding asset is its people, who are articulate, adventurous, canny and mostly cheerful. Under the British Raj, thousands of them were transported to settle as reliable colonists in a swath of places from East Africa in the West to Fiji in the East. It is no accident that Mahatma Gandhi picked up his Indian nationalism while working as a lawyer in Durban, South Africa; V. S. Naipal has explained, in relation to the Indian settlement in Jamaica, how the isolation of the colony there, too small to be concerned with caste or sect, powerfully fostered his sense of being Indian. The legend of Gandhi and the Independence Movement he founded, when the first cohort of political leaders was qualified for office by the time it had spent in British jails, may yet do the same for India itself.



India — a population of over 850 million and comparable to the United States in land area.

Meanwhile, Indians and especially the young have a flair for embracing modernity. It is not simply the designer jeans and the motorized scooters dangerously crowding every city street, but the enthusiasm of young people for science and technology. Can there be an intending bridegroom advertising in the newspapers for a bride who does not boast of a PhD in electrical engineering (and sometimes a 'green card' from the US Immigration Service), and who does not want to marry somebody at least as well qualified, but preferably with an MD? Moreover, as the continuing export of young Indian engineers shows (see page 618), the enthusiasm can survive even the frustrations of student life (but see page 617).

That is the good news. What follows is a survey of the state of science in India, a sequel to that published almost exactly ten years ago (see *Nature* **308**, 581–600; 1984). Much has changed since then. Forone thing, India is visibly more prosperous; the poor may be just as poor, but the middle-classes are more in evidence, probably because they have been enlarged. The bureaucracy, however, is less ostentatious. In New Delhi, there are no longer retainers outside the doors of government officials waiting to fetch and carry for invariably male superiors; many grand people now answer their own telephones. There are personal computers everywhere, mostly made in India. And these days it is often possible to telephone distant cities without having to shout into the handset.

The importance attached to science in India goes back to independence. In part, it stems from the personal friendship of Jawaharlal Nehru, the first prime minister of India, and P.M.S. (later Lord) Blackett, both now dead, who had overlapped at Cambridge in the 1920s. Consistently since then, India has set its sights on becoming a first-rank industrial power on the strength of its own research and development.

What follows will show that there are

many ways in which India has succeeded excellently; the most obvious is the Green Revolution, by means of which the use of improved seeds, fertilizers and irrigation water has allowed India to feed itself comfortably. This year, with the fifth consecutive good monsoon, will bring another excellent crop of grain (wheat, rice and pulses) when the current growing season is over. Moreover, the consequences of a shortfall following the failure of the monsoon can now be calculated in US dollars; 10 per cent less wheat means that an extra $1 billion must be found, no longer a sum beyond the capacity of the Reserve Bank of India. At least for the time being, the threat of famine has been banished. Agricultural scientists such as M. S. Swaminathan are national heroes. (Swaminathan himself, now retired, has set up his own research foundation in a new building south of Madras on the Indian Ocean.)

There is not such a happy tale to tell on population, which continues to grow too quickly for anybody's comfort. The total number of extra mouths to feed is now increasing by some 2.12 per cent a year. This may be less than the growth rate of 2.22 per cent a year in 1980, but it is still so great that, when added to the price inflation rate, it nearly accounts for the whole of India's extra production (measured as gross domestic product) each year. That is one reason why politicians repeatedly wring their hands over the problem, which in a strict sense is eating away at India's seed corn — the extra productive capacity generated each year.

But it is also widely acknowledged that the mere technology of contraception will not bring the birth rate within acceptable bounds, while the example of the state of Kerala (in southwest India) is a vivid proof that the factors that determine low fertility are low infant mortality, good public health and education, even if only at the primary level, especially for girls (see *Nature* **366**, 403; 1993). Yet in much of rural India, young children are not allowed to attend school by their parents, who need the child to help with the animals and in the fields or, otherwise, to earn the Rs4.00 a day's child labour can bring a household. Those who live in urban slums are even less likely to find their way to school.

Prime Minister P. V. Narasimha Rao: tentative steps towards a market economy.

The preferred remedy would be that the public and voluntary agencies seeking to improve the lot of the poor should have time to reach the objectives they have set themselves, but that may not be possible; it is unthinkable that India, with all that pride, should embark on the next century encumbered by a growth rate too great for comfort. The time is coming when politicians, at the union and state levels, will have to risk unpopularity by deciding that the goal of universal education should be enforced. It would be no bad thing if they sought to ease their difficulty by paying parents Rs4.00 a day for every day each child spends at school. If India's dreams come true, after all, it will need a welfare programme one of these days.

The research community has long been aware of what needs doing. Since Mahatma Gandhi, the Indian village has had the reputation of being the most fulfilling form of social organization. So India is proud that there are still 700,000 villages, some of them just hamlets. Some are isolated, others separated from their neighbours only by a couple of kilometres. In many parts of India, the Green Revolution brought irrigation wells (called 'tube wells') and the pumps for extracting water from them. In the more prosperous parts, there are now tractors alongside the buffalo carts. Mechanics and electricians have emerged to service them.

Since independence, great ingenuity has been and is still being lavished on the villages, often by the leaders of the research community. Vikram Surrabai, Homi Bhabha's successor as chairman of the Atomic Energy Commission in 1966, gave this movement a powerful push, partly by his encouragement of the SITE project to project television signals via Earth satellites to single village television receivers. (At the outset, the objective was to spread the word about the use of fertilizers and the new strains of wheat that made the Green Revolution possible.) But now there is hardly a technical agency that does not arrange its technical development to do something special for the villages. Ingenuity is not in short supply, nor are the Gandhian sentiments that cause it to be lavished on the villages.

But India has not yet faced a question that it may have to answer sooner than it would like: is the village the form of social organization best suited to its ambitions for the future? It may on the contrary be the case that the village is the biggest single obstacle to the eradication of poverty in India. The tide of humanity camping in the tin and cardboard shanty-towns on the outskirts of the major cities is made up of people displaced from the villages by the subdivision of family holdings to below the point of viability. The obvious difficulty is that many villages are too small easily to allow for a comfortable division of labour between farming and, say, construction, let alone teaching or health care. And despite the government's traditional encouragement of rural industries such as hand-loom weaving, paid and productive employment is scant. India will in due course have to manage the resettlement of the villages and the amalgamation of their land holdings, with all the social and economic costs that will be entailed. Biting the bullet sooner rather than later would be beneficial.

Will economic reform help? The distinctive policy of the present government of Mr P. V. Narasimha Rao is that it has taken the first steps towards economic liberalization, the standard euphemism for 'market economy' (in which most Indians profess distrust). Since independence, India has been run on socialist principles; major industries and financial institutions such as banks have been (and still are) publicly owned, licences have been required both for imports (for fear that always scarce foreign exchange would be dissipated) and exports (in case the goods might be used at home instead). There have been restrictions on the overseas ownership of Indian assets and restrictions on the freedom with which minority shareholders in Indian businesses have been able to convert their rupee profits into currencies that could be used elsewhere.

Much of this regulatory apparatus is now being dismantled, with beneficial results already apparent. Foreign invest-

Acknowledgements. For the organization of a journey north and south from New Delhi, *Nature* extends its warm thanks to Mr Ashok Parthasarathi, Additional Secretary at the Ministry of Science and Technology, the members of his staff R. R. Abhyankar and Sunrata Banerjee, who were helpful guides on different parts of the journey, as well as the guest houses of the Semiconductor Complex (Chandigarh), the Centre for Cellular and Molecular Biology (Hyderabad) and the Indian Institute of Science (Bangalore) for hospitality and (in the last case) for minor surgical repairs.

Nature also wishes to thank the directors and staffs of numerous academic and industrial laboratories for their willingness to talk openly about their current programmes and problems, as well as for enlightening conversations on more general questions.

This survey has been written by John Maddox, with contributions as indicated from *Nature*'s New Delhi correspondent, K. S. Jayaraman.

Water and employment for all?

LIKE those in what used to be the Soviet Union, India's policies are driven by a sucession of five-year plans. But the eighth plan (1992–97), now in force, differs from its predecessors in its open declaration that its purpose is more indicative than prescriptive. That reflects India's move away from central control of the economy and much else.

But the language of this change is interesting. In his introduction to the latest plan, the Prime Minister, Mr P. V. Narasimha Rao (who is chairman of the Planning Commission and of much else) recognizes that there are many areas in which development can best be ensured by freeing them of unecessary controls", but then insists that India's growth and development cannot be left entirely to the market mechanism. The market, he says, may be able to balance supply and demand, but not supply and need; planning is essential for taking care of the poor and downtrodden.

There will be plenty of those. The plan says frankly that if the impoverishment caused by rapid population growth is not halted, it will never be possible to render social and economic justice to millions of our masses. The planners accept received wisdom that this will have to be done by education, health services (to reduce mortality) and by provision of paid employment. (Ambitiously, they make full employment by the end of the century one of their goals.)

For the rest, the objectives for the current five-year plan are the provision of safe drinking water for everybody (electrification of the villages was the target of the two previous five-year plans), the diversification of agriculture (partly for the sake of the exports it will generate) and the improvement of the infrastructure of the growing cities (in which 25 per cent of Indians now live).

The new plan pencils in a population growth rate of 1.87 per cent a year in 1992–97, suggests that a net reproduction rate of unity (which implies mere replacement of the population) would be attained roughly a quarter of a century from now, but that, because of the predominance of young people in the population then, numerical stability will be reached only towards the end of the next century. Prudently, the planners do not say what the stable figure will be.

On science and technology during the five-year period, the plan says that attention should be concentrated on projects contributing to its declared social goals as well as to the improvement of industrial production. But the plan also declares that research projects that would give India an international reputation should be backed, as should be the projects (such as India's interest in space) already accepted as strategic objectives.

ment in India is growing quickly and overseas trade is picking up. The snag is that it has been necessary to protect the domestic economy by fiscal rather than physical means: the rupee has been devalued (to Rs31.00 to the US dollar) but is convertible, the banks charge 18 per cent interest on commercial loans and government spending is being kept on a tight rein. Businessmen delight at the erosion of the licensing system and at the way in which that has robbed pettifogging bureaucrats of the power they previously enjoyed, but complain that foreigners are snapping up Indian assets too cheaply because interest rates in India are so much higher than elsewhere in the world.

The Rao government must now decide whether to persist with its reforms and, in particular, whether to begin denationalizing the major industries and the banks. The state assembly elections in Northern India at the end of last month will not have helped to clear its mind: both the government party, the secular Congress(I) (where TIU stands for 'Indira'), and its chief rival, the BJP (a Hindu party) did less well than they might have done, but the government was not so severely punished that it will be compelled to count its economic reforms mistaken.

Yet what the government does about economic reform in the next two years (before the next general election) will be crucial to India's future, and to that of Indian science. The only known way of quickly generating the resources required for social reconstruction is the market economy, but dashing for full economic liberalization will bring a painful social upheaval in its own right. The Rao government is in much the same cleft stick as many of those of Central Europe, faced with the task of unwinding a command economy without damaging the social fabric irreperably.

What will the reforms do for Indian science? The chief effects so far have been painful and unwelcome: frozen and even shrinking budgets caused directly by the need to keep government spending under control. Universities and basic research institutes are the chief victims. University budgets (in rupee terms, without allowance for inflation) have been frozen for two years and there is no relief in sight. And many research laboratories are now being required to arrange that at least half of what they spend each year will come from sources other than their sponsoring agency by 1997. As well as the reduction of public spending that will result, one objective is to persuade Indian companies to take research more seriously. The snag is that some laboratories are incapable of feats such as these, and that the new policy represents a sharp break with the post independence tradition that the 'centre' (the collective name for the union government and its agencies) decides and carries out the research that India needs.

The reforms thus require decentralization of decisions on research, for which the machinery has fortunately been constructed in the past few years. Indeed, there have been striking changes in the past decade. Laboratory directors unable to make simple decisions because they had (and wished) to refer to Delhi are now far fewer. But it remains to be seen whether the public laboratories (and the government agencies that support them) have entirely shaken off their old ways of working, the effect of which was commonly to allow ambitious projects to continue beyond the point of vanishing success without responsibility being attributable to anybody. The most serious danger in present circumstances is that attempts to create in India a replica of one of the many, and mostly false, replicas of Japan's Ministry of International Trade and Industry will prove to be another scheme for enabling the centre to back research projects of its own fancy, Twinners' and, as often, losers.

In passing, the parallel between the present problems of the research enterprise in India and in Britain cannot fail to capture the attention. For different reasons (the logic of economic reform and impatience respectively), both governments are seeking closer links between the research they support and industry more generally. In each case, researchers are required to take the initiative, winning first the interest of industrial partners and then, with luck, their financial support. If anything, India's problems are the simpler. Maybe on this occasion, Britain has more to learn from India than the other way round.

Where the effect of India's economic reforms on the pattern of research will in the long run be most marked is in its policy of 'self-reliance'. Self-suffiency (which means 'self-contained') is no longer a credible, let alone a desirable, goal. But India has taken from the outset in 1947 the view that it must be able to rely on its own resources for the technological skills it considers vital to its survival. That explains the energy put behind the engineering of the Green Revolution, which has rid the government of the need to beg annually for credits with which to avoid famine somewhere in the country, and in the process to be preached at by Western politicians on matters as different as relations with the Soviet Union (close

while the Soviet Union was still extant) and the domestic rate of population growth.

That also explains the existence in India of the Atomic Energy Commission (AEC), originally a means of security on power supplies, but latterly also the probable custodian of a small stock of nuclear weapons. (When visitors say to knowledgeable scientists something like, 'Since that Tpeaceful nuclear explosion' in 1974, the rest of us have assumed that India is already a nuclear power', national pride drives a smile to their faces before they reply that they assume so too, but do not know for sure.) Historically, the AEC was followed by the Electronics Council (from which has sprung many of India's present investments in research and development) and by the Indian space programme. If the future hangs on microchips, being able to make them is an essential guarantor of strategic, political and economic independence.

But why not follow a middle route, buying licences to make, say, chips from Intel and then following South Korea towards prosperity? India will have none of that. Too many compromises would be necessary. People seem not yet to have tumbled to the truth that it is also a demeaning compromise to have more than a third of the whole population below the officially defined poverty line, and to see no obvious remedy in sight.

The plain truth is that India is still fighting in its main the old battle of the occupation by the British Raj, and is willing to pay a high price for the freedom that victory will bring. When it appreciates that the battle was won half a century ago, it will be readier to concentrate on what it does best, as an economic division of labour will eventually require. The test of India success will be its willingness to shade the goal (or shed the burden) of self-reliance.

In short, India is now well placed to bargain its way, as an equal partner, in many constructive technical projects. The chemical industry is already working in this direction, but the opportunities in electronics are just as attractive. That may be one of the benefits of the economic liberalization now under way.

Part of the difficulty, which shows up as diffidence, is that India seems not yet to appreciate how much it has achieved in the past ten years. Starting from the top, the administration of the research enterprise has been made much less bureaucratic than it used to be. Of course, it would be ridiculous to suggest that all decisions are still made on objective grounds, divorced from political and other considerations, but there has been a great improvement.

The glaring need is for some means, at the higher reaches of policy-making, for a mechanism for comparing the value to India of its investments in different fields. The scale of India's investment in science and technology also needs attention. By some counts, there are more than a million people working as scientists and technologists. But given the scale of India and of its social problems, that is not self-evidently too great a number. On the contrary, if India is ever to realize its post-independence dream of prosperity through science and technology, even larger numbers of people will be needed. That is why the long-standing neglect of the universities, recently intensified, is a serious black-spot in the present state of affairs.

The result of the deprivation of the universities has been that main-stream researchers have gravitated to government laboratories and the government-supported independent institutes, many of which now have facilities and a sense of where they are heading which is in every way comparable with those elsewhere.

Indeed, in many of these laboratories, the sense of mission can only be accentuated by what researchers see every day on the outskirts of their cities – the shanty-towns and hamlets that provide a simulation of shelter to India's poor. Is it not at once sobering and inspiring that a seemingly abstract project to determine the antigenically effective epitopes in a mycobacterium, or to strengthen the human immune defence against the rotaviruses (responsible for intenstinal infection) may one day, perhaps soon, help those poor people to survive? Where else can there be such a clear promise that laboratory work will so immediately improve the condition of fellow citizens? Inevitably, of course, not all public laboratories have this sense of élan. But the administrators now seem seized of the need that unproductive laboratories should be reorganized, and that people who are no longer individually productive should somehow be shunted in other directions. But even within the circle of alert laboratories, obvious improvements are needed.

Thus India underestimates the isolation of its researchers, not only from colleagues overseas, but domestically. The research enterprise would be more effective if there were a more ready exchange of people within the public system, and between public laboratories and the universities. Laboratory directors meet each other at committee meetings, but less exalted people have only a diminished sense of what their colleagues elsewhere in India are about.

The case of India's present love-affair with computers and communications is nevertheless an object-lesson in success, in this case attributable to the enthusiasm of a few individuals. The origins of the programme are twofold—the recognition that, for geographical reasons, satellite communications would have an important part to play in India and the calculation that India was potentially a competitive manufacturer of electronic equipment. But at the outset, in the early 1970s, attempts to provide the villages with improving television programmes seemed improvised and amateurish.

Although the hope of giving every Indian village a sense of more general community may not yet have been realized, the remarkable outcome of this earlier enthusiasm has been an extraordinary flowering of schemes as different as those for warning of typhoons in the Bay of Bengal and for storing nationwide administrative information in a single database (see page 622). India itself does not appreciate the interest of these developments, which are largely unrecognized elsewhere.

Why have developments such as these not given the Indian research community greater confidence in its capacity to engineer beneficent change? Because that takes time, and because it requires examples of how Indian science can perform well even in fields regarded elsewhere as high points in basic research. That is the strategic function of the independent research institutes scattered through India, as well as of government laboratories such as the Centre for Cell and Molecular Biology at Hyderabad.

The research community appears to recognize recent achievements of this kind. Its temper has changed remarkably in ten years. People have a more realistic assessment of what they can achieve with the resources at their disposal, and of the external competition. There are many fewer tendencies to believe that a few badly typed sheets of paper can contain a proof that Einstein was mistaken about special relativity, or that human consciousness can be described by an appropriate field theory.

So how to explain abberations such as Professor V. J. Gupta's distortions of the Himalayan fossil record?

Sober opinion in India is either too embarrassed to talk about the problem, or so conscious of the legal difficulties facing the Panjab University at Chandigarh that it shrugs its shoulders. But that is merely a pointer to the need that India's universities should enjoy a greater sense of freedom than they have.

The paradox, for India, is that its justifiable pride in what the research community has accomplished in the past ten years has not been matched by its trust in the community as a whole. The suspicion that the present government is less persuaded than its predecessors of the importance of the scientific enterprise is especially worrying. It is to be hoped that the coming decade will exorcise that for good. □

How India runs its science

THE machinery of the government of science in India is partly a legacy of the British Raj. There is still, for example, a Department of Scientific and Industrial Research (abolished in Britain in 1963) and a National Research Development Corporation (NRDC) for the exploitation of publicly generated intellectual property; its British counterpart was first renamed the British Technology Group and then privatized, earlier this year. But the most powerful agents of Indian science are the newly created operating agencies controlling semi-industrial sectors of the technical economy such as atomic energy and space research.

There is a clutch of government ministries with technical responsibilities, of which the most obvious, but not necessarily the most influential, is the Ministry of Science and Technology. Technically, its head is the Prime Minister, who holds the cabinet portfolio for science and technology. But day-to-day business is the responsibility of a Minister of State for Science and Technology, Mr P. R. Kumaramangalam.

Kumaramangalam made something of a stir in the wake of the state assembly elections last month by writing to the prime minister (and releasing to the press) a letter complaining that their party would have done better if its leader (the prime minister) had more decisively adopted policies previously suggested. He went too far; on 2 December he resigned.

The ministry is responsible for three operating agencies – the Department for Science and Technology (DST), the DSIR and the Department of Biotechnology (called DBT). DST has the responsibility of making policy and of promoting science and technology more generally. Thus it makes grants for research projects (through a Council for Science and Engineering Research), supports professional associations and societies — including India's three rival science academies, the Indian National Science Academy (New Delhi), the Indian Academy of Sciences (Bangalore) and the National Academy of Sciences (Allahabad) — and runs projects as different as the Geological Survey of India and a scheme for the avoidance of disasters from typhoons. DST is also the chief source of support for nationally important research institutes such as the Raman Institute (Bangalore) and the Bos Institute (Calcutta).

DSIR, radically reorganized in 1985, is a technology transfer agency (which thereby controls the NRDC) but which also contains a Council for Scientific and Induatrial Research (CSIR). That agency runs a network of 29 research laboratories as different as the Building Research Institute at Roorkie (north of Delhi) and the Centre for Cellular and Molecular Biology at Hyderabad, which is not easily distinguished from a basic research laboratory.

The Department of Biotechnology, created in 1984, is a monument to India's conviction (and in particular to that of Dr S. Varadarajan, then chief secretary to the ministry) that there were important opportunities in this emergent field. The department has two laboratories of its own, including the excellent Institute of Immunology at New Delhi, but also controls a number of state-owned companies, including the Indian Vaccines Corporation Ltd (whose chairman in Varadarajan) for manufacturing viral vaccines.

In strictly money terms, other ministries may have a larger influence on India's civil science. These include the departments responsible for atomic energy, electronics, space, ocean development and chemicals and fertilizers.

Another formal source of funds for civil research is the University Grants Commission (UGC), which is technically a part of the Ministry of Human Resource Development. (Responsibility for education in India is shared between the union government and the governments of the 25 states that make up the union, but certain institutions have been singled out as being of national importance and are directly financed from the centre.)

The UGC's subventions to the more than 150 universities in India are supposed to provide academics with the infrastructure of research support. But the intended recipients say that the UGC has failed in recent years to support them adequately. The explanation has much to do with the vast scale of India's university system. There are more than a hundred universities and comparable institutions, and more than 4 million students in higher education.

Responsibility for most undergraduate education is delegated to the 4,000 or so four-year colleges scattered throughout India, many of them missionary foundations in the eighteenth and nineteenth centuries. Curricula are determined by the universities, which also provide the examinations that must be passed before graduation. Ambitious graduates usually follow a two-year masters' degree course at the university proper. The better universities maintain PhD programmes, but increasingly graduate students in the sciences elect to be based at one or other of India's free-standing research institutes.

Among universities, six are exceptional. There are five Indian Institutes of Technology (IITs), founded in the 1960s with assistance from well-wishing governments overseas, which have established themselves as élite engineering schools. There is also, at Bangalore, the Indian Institute of Science, founded late in the nineteenth century, which has become India's near-analogue of the Massachusetts Institute of Technology.

In principle, India thus enjoys a delicious plurality of sources of research funds. All ministries with a technical brief are able to make research contracts and grants. Successful institutions, may at any time benefit from research contracts with many different government agencies.

That is the pinnacle of succcess; other universities complain that there is just too little money for their research needs and that the procedures for finding it are not nearly as systematic as they should be.

Traditional links between ministries and research institutes they have helped to found account for a good deal of the flow of funds. Thus the outstanding Tata Institute for Fundamental Research (TIFR) at Bombay, founded by Homi Bhabha with a benefaction from the Tata family and consistently supported by the Atomic Energy Commission, is now the prime mover (and chief source of funds) for India's ambitious programme in radioastronomy (see page 620).

In ten years, much has been done to decentralize decision-making. Laboratory directors can now hope to be their own men (there appear to be no women yet) provided that they can live peacably with the scientific councils, composed of academics and potential customers, that meet about three times a year. The departments themselves are also attended by advisory councils.

What most worries working scientists is the shortage of funds. Among government agencies, defence (27.5 per cent), space (15.5 per cent) and agriculture (13.1 per cent) consume more than half the total, which is in any case a declining share of the gross domestic product. (India's research spending has fallen from its peak three or four years ago to 0.9 per cent of GDP). Some, such as C. N. R. Rao, the influential director of the Indian Institute of Science, see this as proof of the government's diminished commitment to science.

One glaring lack is that of a central advisory committee. The present prime minister has not followed his predecessors in appointing a chief scientific adviser and a powerful committee to give high-level advice. Previous arrangements of this kind were not all that successful, but especially at a time of rapid change such as this, Rao is much in need of a device for making sure that the blunt instrument of reform does not irreparably damage existing institutions, among which the universities are the most vulnerable. □

Palaeontology under a Himalayan shadow

CHANDIGARH is the northern city now belonging to no state of India, but instead sandwiched between the Punjab and Haryana, respectively Sikh and Hindu predominantly. It is also the city that the French architect Le Corbusier sought in the 1930s to turn into an architectural paradise with several striking buildings; although they have not weathered well, their presence seems to have given the city as a whole an architectural self-consciousness not apparent elsewhere.

But Chandigarh is also the home of V. J. Gupta, professor of palaeontology in the department of geology at the Panjab University of Chandigarh. Gupta is now best known not for the significance of his Himalyan fossil finds, but for the allegation (see J. A. Talent, *Nature* **338**, 613; 1989) that many of his fossils, described in more than 200 publications, do not derive from the Himalayas but from where in the world, often the palaeontological collections of university departments that Gupta had visited on his travels.

What follows is an account of a conversation with Gupta, as well as with others at the university concerned in the affair. It is proving to be a protracted business. Talent's accusations that Gupta had systematically salted Himalayan rocks with specimens from where are now six years old. Since then, the accusations have been endorsed by two independent investigations — by the Indian Geological Survey and by an expedition to the supposed sites of the fossil finds under the leadership of Dr A. S. Paintal. Gupta was at first suspended from the university, but then reinstated. For the past two and a half years, he and others have been giving evidence at an enquiry conducted by a retired judge, M. S. Gujrat, which is now in its closing stages. There should be a judgement during 1994.

Since Talent's article appeared, there has been some concern that some of Gupta's associates at Chandigarh, notably Dr A. D. Ahluwalia and Professor S. B. Bhatia, appeared over-splenetic in their denunciation of Gupta, with whom they had jointly published articles. A conversation over lunch in Chandigarh made it crystal clear that each has been deeply wronged by Gupta. Bhatia, being the older, is the more philosophical.

To the complaint of colleagues in the geology department that they were happy enough to go along with Gupta while he was able to publish articles with their names included, and that they have turned against him to save their own skins as potentially culpable co-authors, they reply that they have no choice. Bhatia is nearing retirement while Ahluwalia, who does not think he could easily find a comparable job elsewhere, is fighting for his livelihood.

V. J. Gupta himself is more urbane. On the telephone when agreeing to this interview, he had explained that he would probably not be able to discuss the details of the case because there is a judicial inquiry going on. In the event, in the comfort of his house in a row of three, at the bottom of the campus and armed with a stack of papers by his side, he was more than ready to discuss the details of the case.

Why had *Nature* published Talent's accusations without letting him, Gupta, see them in advance? Was not this or that accusation, that he could not have visited some of the sites recorded in the published papers, contradicted by this piece of paper in his hand consisting of a *laisser-passer* from the military authorities for the locality concerned? (One of the difficulties is that many of the Gupta sites are in India's disputed border regions with China.)

For fear that this kind of argument would easily occupy the available hour, it seemed best to get down to basics. So is it not the case that the judicial inquiry, whatever its outcome, is irrelevant to his reputation in the scientific community? Is it not the case that the scientific community believes Talent, not Gupta? Did Gupta believe he could rehabilitate his reputation in the face of general disbelief?

Disarmingly, Gupta agreed that his reputation is in tatters. Implying confidence that he would be exonerated by the investigation under way, he also acknowledged that it would be difficult to win back his reputation among palaeontologists even when the inquiry was complete. But he would do whatever was necessary.

What about giving people elsewhere full access to his fossil specimens? Even though they are mostly microscopic entities, it might just be possible to identify some features of the rock in which they had originally been embedded?

Gupta's reply to that was breathtaking. "But I have some specimens of rock that have not been ground up. I've not told people about this. I've kept them to myself." The surprise in that, of course, is that one such as Gupta now exposed to the third inquiry into his integrity as a scientist should have neglected to disclose to any of them evidence that could well establish his innocence of the charges against him. That he should have retained some unprocessed rock is natural, but that he should not have used it in his defence is mystifying.

In an isolated city a long way north of Delhi, it is tempting to sympathize with the co-authors' alarm. Will it be, a few months from now, that the remaining rock will be ground up, found innocent of fossils (conodonts in particular) and the coauthors then accused of having spiked the original samples with material from elsewhere than the Himalyas?

Later that same evening, it emerged from a conversation with Dr R. P. Bambah, vice-chancellor at Chandigarh until mid-1992, that there are other more mundane complications. One of them concerns the status of Indian academics, which is that of government employees, who cannot be deprived of the privileges of the posts to which they have been appointed without prescribed due process. Bambah says that, without an independent quasijudicial inquiry in the form that he established, any move against Gupta parmanently to attenuate his academic privileges would have been overturned by the courts.

Back in Delhi, Paintal seems to fluctuate between annoyance and resignation, impatient of the legal niceties of the Gupta case. But his report (supplied by another source) is not quite the weighty document the circumstances would seem to require, consisting of a one and a half page covering letter enclosing laboratory notes of investigations by specialists on the expedition to Spiti, in the Himalayas.

The practical consequences for Chandigarh of the Gupta affair are so far two. First, the University Grants Commission has deprived the geology department of its special status, which those opposed to Gupta say has hurt only the students there, while the geology department seems to spend its non-teaching time mulling over the details of this murky affair. Personal animosity (on both sides) is palpable.

Whatever the outcome of the judicial enquiry, it will be irrelevant outside Chandigarh. Gupta's confidence that he can rehabilitate his reputation betokens, at best, a failure to appreciate the seriousness of the misdemeanours embodied in the documentary evidence in the literature, at worst a belief that the research community's attention span is short.

In Chandigarh, the verdict will matter a great deal. The atmosphere in the geology department is poisoned. Honest coauthors have been castigated for not having complained earlier, and then their complaints have been described as attempts to save their own skins. But if the enquiry finds no charges proved, Gupta will be next-but-one in line to be Dean. The university cannot afford to wait for the judicial enquiry to end before tackling those problems. □

Higher education with lower budgets

INDIA'S science has a great and probably unique asset: the enthusiasm of young people for science and technology. It is a remarkable phenomenon. India, always a society obsessed with matters of the mind, values academic qualifications as tangible proofs of intellectual attainment. But qualifications are also routes to employment, and jobs in science and technology have traditionally been especially sought after. (Only now are some academics shaking their heads over the extra status attached to an MBA.) So the competition for entry to science and engineering is fierce.

So much can be told from the competitions run by the Indian Institutes of Technology (IITs), and by the Indian Institute of Science (among first graduates from other universities). Entrants to the competitions run by the IITs routinely outnumber the places available by factors of up to 1,000. So it is a cruel disappointment that many universities (but not the IITs or the Indian Institute of Science) say they are now unable to teach their students well for lack of equipment and facilities.

The complaints of the physics and astrophysics department at the University of Delhi (see *Nature* **366**, 297; 1993) are typical. Academics worry about several pressures on the system, one of which is said to be a decline of the quality of teachers at the constituent colleges at which entering students read for their first degrees. That, the argument goes, means that the chance that students will be freed from the burden of rote learning is bound to be much diminished.

And why should that be? Because the academic life is full of frustration, and badly paid as well. On both counts, shortages of funds are the root cause. Thus in the financial year (to the end of March) 1991–92, the union and state governments were expected to spend Rs3.26 billion on direct support of 177 university institutions (with research funds from technically orientated agencies routed separately). But with 4.4 million students at universities in that year, the support works out at just over US$20 per student a year.

For what it is worth, the 1992–97 plan supposes that there will be an extra one million students at India's universities by the end of the period, but that half of them will have to be taught by the several open (or 'distance-learning') universities being founded in India. That development will be a serious test of young people's eagerness to acquire academic qualifications.

In the circumstances, it is hardly surprising that shabbiness is the hallmark of the regular Indian university. Classrooms are much as they were when first built, perhaps in the early decades of this century. When university libraries are agonizing over the need steadily to reduce their purchases of books and journals, who can wonder that paintwork is allowed to flake off, or that classrooms and offices accumulate dirt faster than it is cleaned away?

The shortage of equipment drives people to despair. Departmental budgets for maintenance and renewal are negligible, consisting mostly of the fees (around Rs20) that graduate students pay each term. It is no wonder (this is the complaint at Delhi) that academics say they are ashamed at being unable to do their duty by their students.

Yet academics, however their self-respect is undermined, are still respected. Morover, by Indian standards they are reasonably well rewarded; a junior member of a departmental staff may earn the equivalent of US$75 a month, a full professor twice as much. In addition, a university has a duty to provide either housing of a kind or (where accommodation is to be had commercially) a housing allowance.

The last of these circumstances contributes to the low mobility of academics in India. Indeed, despite the requirement that vacancies for public employees (as academics are) should be open to general competition, there is a natural tendency for people to stay where they are. That no doubt explains why many (but, emphatically, not all) departments at universities in India have the air of being groups of people united only by their mutual suspicion, even jealousy.

In the circumstances, there can be very little comparison between the quality of the scientific life at the harassed Indian universities and at the better public research laboratories, where the buildings are often new, equipment is plentiful and access to journals (and even to international computer networks) almost routine. Especially because the research laboratories are increasingly welcoming of PhD students, it seems inevitable that the universities will be further starved of able researchers.

On the face of things, the years ahead will be even more depressing. Budgets are being frozen in rupee terms, with no allowance for inflation and little prospect of relief. Even at the Indian Institute of Science, people are alarmed at the signals emanating from Delhi.

The government is wanting us to stand on our own feet, says Dr G. Padmanaban, deputy director and professor of biochemistry at the institute. In the past few years, the institute has indeed opened up two new lines of activity – continuing education for people with a technical background and the engagement of members of its faculty in the provision of consultancy to industrial organizations, publicly and privately owned.

The institute earns Rs2 million from the first of these activities, perhaps ten times as much from the second. But its directors say they would be reluctant to expand the work of the Centre for Scientific and Industrial Consultancy to the point at which members of its staff were engaged on routine consultancy lacking intellectual challenge.

Elsewhere in the academic sector, there are persistent rumours that the Delhi government will seek to come to terms with the high rate of emigration from the Indian Institutes of Technology (see page 618) by raising student fees to levels at which the institutions were almost self-supporting financially. What that will do to recruitment remains to be seen.

Outsiders cannot but be sympathetic of the problem of higher education in India. Even the problem of the Russian universities is more easily manageable. But there is a serious danger, in India, that the national ambition to win prosperity through indigenous technical development will be undermined by the steady erosion of quality in the universities.

What is to be done? Given that the government could not at present afford to support all 177 universities in a manner that would enable them to function well, it could at least single out a number for special treatment in the expectation that the outcome would be a group of institutions capable of serving as an example, or at least a goal, to the remainder. The difficulty is that such a course was followed several years ago, when a handful of universities was chosen for just such reasons. (The University of Delhi was one of those.)

Now there seems no alternative to going through the same procedure all over again. But on this occasion the experiment might be more radical. The effective separation of undergraduate from graduate teaching by means of the system of undergraduate colleges may be a way of dealing with large student numbers, but otherwise is a thief of young people's time and energy.

The colleges may be necessary to put human faces in front of students, but their sharp separation from the universities at which their students graduate hardly seems necessary.

It is also important that the universities should enjoy a greater degree of autonomy. It should not be as difficult as at present to reorganize the curriculum or the disposition of teachers' time. And too often, decision-making consists of guessing what the state governments are likely to decide next, which is debilitating and demeaning. □

An unwelcome export success

Is India now one of the world's chief sources of migratory technical skill? According to two recent studies, three out of ten engineering graduates produced by the Indian Institute of Technology (IIT) Bombay since 1970 have settled abroad, while the brain drain from IIT Madras varied from 20 per cent during 1963–67 to 35 per cent in 1983–87. The five IITs are the country's most distinguished engineering schools.

Things are much the same at all five IITs. Thus Professor M. C. Nigam, director of IIT Delhi, estimates that the exodus from his institute has been steady at 30 per cent. And it is generally agreed that nearly a quarter of the engineers trained at all the five IITs (the others are at Kanpur and Karagpur) take a plane to the West each year.

Computer scientists head the list. In a recent study, Robert K. Perkins of the University of California, Berkeley, found that 91 per cent of first-year computer science students of IITs and the Jadhavpur University planned to emigrate when they were qualified. In fact, 70 per cent of them at Jadhavpur ended up going abroad.

The situation is much the same in medicine. A recent study by Professor Veena Kalra of the All-India Institute of Medical Sciences (AIIMS) in New Delhi discovered that 45 per cent of graduates from the institute since 1971 have emigrated. She says that the brain drain, which reached a peak at 80 per cent ten years ago, continues but at a lower rate.

"Considering that it costs taxpayers £3,000 to educate one IIT graduate and twice that amount for a medical student, India has repaid through export of human capital more than the total aid it received from abroad," says S. Biswas, an education consultant in Delhi. "India, no doubt, produces more professionals than it needs," he says, "but the 2 per cent of professionals who leave each year represent the cream."

Meanwhile there are fears that the brain drain will accelerate because of the recent economic reforms, which encourage the entry of multinational corporations into India. "When multinationals come to India in a big way, they will draw on local talent and, after a few years, relocate them to Singapore, Korea or the United States," says Professor Nigam.

While the migration continues, its pattern is changing. Thus there has been a recent growth of migration towards Australia, which has opened its doors to technically qualified people, and to Singapore, which has begun giving work permits to Indians. Nigam believes that the brain drain will continue unless Indian industries modernize and unless the government offers world-class facilities and opportunities.

One feature of the migration from India now apparent is that people seize the earliest opportunities to leave. Thus the bulk of the exodus from India to the United States is made up of fresh graduates from IITs and other institutions of higher learning. Once people are established in posts at universities, the Council of Scientific and Industrial Research (CSIR) or the councils of medical or agricultural research, the risk that they will migrate appears to be negligible.

Past schemes to stem migration appear to have had little success. Indeed, the Indian government does not seem to be unduly worried; its new technology policy statement offers no prescription for reducing the brain drain.

"The brain drain is a natural process and cannot be plugged", says S. C. Mazumdar of CSIR, who is in charge of the TOKTEN (transfer of knowledge through expatriate nationals) project to reverse the brain drain. But in 13 years, he has been able to tempt only 500 expatriates to accept short-term assignments in Indian laboratories. The project is in any case to be wound up in December 1994. Another government scheme, launched two years ago to lure expatriates to industry, has yet to yield results. The 35-year-old "scientist pool scheme", designed to find jobs for returning scientists, has so far succeeded in resettling about 7,000 Indians, but many have since emigrated again.

Mazumdar nevertheless emphasizes that even when Indian expatriates in the United States do not want to return, they are eager to help in other ways. He mentions, for example, their willingness "unofficially" to arrange for training of their countrymen who happen to be visiting the United States.

Even so, the government's hopes that it would keep able people at home by the creation of new technical agencies, in fields such as ocean development, environment and biotechnology, appear to have been disappointed.

But there are some exceptions. The National Informatics Centre (NIC), which has 3,200 computer professionals, boasts of having lost only 60 in 15 years. In contrast, 15 per cent of the staff of Tata Consultancy Service, a private computer company, have gone abroad despite better pay than those at NIC. "People do not want to leave NIC, even though their paychecks are not thick," says N. Seshagiri, NIC's chief. "We give them intellectually challenging problems in frontier areas of technology, and we have the best equipment and a good personnel and promotion policy. We have proved that brain drain can be stopped under the right conditions," he says.

K.S.J.

A gold mine of information

SELLING Russian science to the West India's National Informatics Centre (NIC) in New Delhi is planning to exploit Russia's need of cash by acting as an information broker.

The calculation is that there is a huge amount of saleable information on science and technology (S&T) which the world is not aware of because it is all in Russian. And several companies in the United States are ready to pay money for translated Russian literature in the belief that it contains ideas or inventions they can exploit commercially. According to N. Seshagiri, NIC's director-general, there is a goldmine of information in theses, laboratory reports, journals and other Russian-language documents that the Russians now wish to sell.

NIC has worked out a two-way arrangement under which it will translate the Russian reports, create appropriate databases and then make these available to vendors in the United States. We are just a middleman, says Seshagiri. From whatever the US companies pay, NIC will subtract a service charge and hand over the rest to the Russians. Seshagiri says it will be a handsome amount.

Four Russian institutions, all in Moscow, have entered into this deal with NIC. They are ICAD (Institute of Computer-Aided Design) and VINITE (a science and technology information organization), and two institutes of the Russian Academy of Sciences. Initially, the agreement will run for two years, but will be extended if there is a sufficient demand for it. NIC has negotiated with three US companies, who will be in Delhi this month to finalize the agreement.

NIC is opening an office in Moscow and installing computer terminals that will be linked to NIC's Cyber–730 host computer. Full translation of Russian texts will be provided on request. To start with, NIC will prepare databases on fluid mechanics, aerodynamics and informatics.

According to Seshagiri, India and Russia will invest 1 million each in this joint venture, which he says will give a return of 10 million annually beginning in 1994. The profits will be shared between Russia and India in the ratio of 60:40.

K.S.J.

A model institution on trial

THE All-India Institute of Medical Research (AAIMR) in New Delhi is now the proud owner of India's first diagnostic NMR (otherwise Magnetic Resonance Imaging, or MRI) machine. One day late last month, the people at the monitoring desk were looking intently at an image of the spine of a tiny woman, even then turning round to get off the narrow bed just emerged from the magnet. Chillingly, they explained that the dark spot on one vertebra, perhaps the size of a dime, was a tuberculous lesion. They said she had been complaining of a pain in her back. India plainly needs more NMR machines.

AIIMR is always an excellent place at which to embark on an enquiry into India's medical research. It is a unique institution, an amalgam of the US National Institutes of Health and a medical school with an annual intake of 50 beginning doctors, an equal number of qualified physicians seeking postgraduate qualifications (usually by research) and as many nurses. Founded in 1956, it was meant to be a model medical institution for the rest of India; the high rate of emigration among its graduates is presumably a testimonial of success.

Despite the NMR machine, it may be that a little of the glitter has been lost at what is still an outstanding institution; oddly, for example, there is no department of medical genetics among thirty-odd academic departments, while a promised review of the curriculum promises to be less radical than recent decades would suggest possible. But the institute maintains a community hospital and a clutch a health centres (at one of which each graduating physician must spend three months of an internship), so that it is an excellent listening post.

So what is at the front of the field-workers' minds? The dean (after gravely shaking his head over population growth) puts it succinctly: tuberculosis (which has always been a problem), malaria (and what a tragedy that the eradication campaign ran into the sand a decade ago!) and virally infectious hepatitis not A, not B, but E. At least for the time being, AIDS is further down the list.

Other institutions endorse by the emphasis of their research this ordering of emergent public health problems. It may seem cruel that India is plagued by two serious diseases caused by mycobacteria — leprosy (*Mycobacterium leprae*) and tuberculosis (*M. tuberculosis*), but that may not be entirely a coincidence. Many good laboratories are looking for better ways of dealing with these organisms.

At the Indian Institute of Science, Bangalore, Professor K. P. Gopinathan has a group working on the basic molecular and cell biology of the mycobacteria. Why, for example, does the cell division of *M. tuberculosis* take the best part of a day, compared with 40 minutes for *E. coli*? Does that slow pace of replication somehow explain the relativity immunity of the TB and leprosy organisms from the usual immune response?

The objective is to understand the molecular biology of mycobacteria well enough to guess which kinds of drugs may be effective. Gopinathan was the first to show that a peroxidase in the mycobacterial cell membrane may be linked with the reasons why the drug isoniazid is lethal to mycobacteria, and also with the reasons why resistance quickly appears. Is that set of puzzles a route to better drugs? Or, having isolated the topoisomerase enzymes that regulate the degree of coiling in the DNA, are there ways of interfering with their vital functions by means of drugs? Rather wistfully, Gopinathan asks whether a complete sequence of *M.tuberculosis* would help, observing merely that a great deal of time and effort would be involved in that.

The National Institute of Immunology (NII), at New Delhi, is following a different route to protection against *M. leprae* – the use of a vaccine based on a distantly related mycobacterium first killed by heat-treatment, on the same principle as the use of vaccinia (cowpox) virus as a prophylactic against smallpox. The vaccine is already in clinical trials at two Indian hospitals, where it is being used in people who are already infected with promising, but not startling, results. When that trial is complete, the next step will be to test its efficacy against unprotected people at high risk.

NII is the creation of the Department of Biotechnology, itself relatively new, but is the brainchild of Professor G.P.Talwar, its director until his retirement in 1991. Among Indian laboratories, NII stands out not only by having an international reputation, but also because half of its income comes from overseas, from the UN Population Council, the US Rockefeller Foundation and Canada's International Development Council among others.

Dr Sandip K. Basu, now the director at NII, is particularly pleased that international recognition brings a steady flux of visitors from overseas, perhaps 25 to 30 a year. The laboratory also takes in 10 or so beginning PhD students a year, recruiting them from all over India.

NII's fame rests on its anti-pregnancy vaccine, now being tested in some hundreds of women volunteers. The principle is to induce an immune reaction against human chorionic gonadotrophin (hCG), the natural hormone involved in the implantation of the fetus. What the research so far has shown is that women with antibodies against hCG capable of neutralizing 50 ng per ml of the natural hormone only rarely become pregnant.

Basu says that it is a long way, perhaps 10 years, from here to a usable vaccine. For one thing, although a relatively simple protein, hCG consists of two peptide strands which are highly glycosylated (meaning that it cannot be easily manufactured by recombinant DNA techniques). The immunity so far established has been brought about by a combination of the human alpha and the bovine beta strands, but a large part of NII's effort is at present going into the search for parts of the intact molecule that will be effective antigens and for cloning vectors that may allow the relatively easy production of the components of the vaccine.

Basu and his colleagues recognize that there is a difficult road ahead of them. A vaccine with several components will not be easily administered in a country such as India, whence (in part) the search for ways of attaching them to biodegradable materials giving slow release over a period of time. There are also important questions about the longevity of protection; too little may make the effort ineffective, too much may mean that the antipregnancy immunity offered by the vaccine is irreversible.

They are also conscious of the difficulties that will arise if and when the anti-pregnancy vaccine is shown to be effective and practicable, when it will be necessary to persuade women to adopt this form of birth control when they may consider that the responsibility lies with their husbands. In the circumstances, it is as well that the institute is also seeking a vaccine to be used in males to prevent the formation of viable spermatozoa as well as barriers to fertilization of the human ovum.

Meanwhile, the hunt for better diagnostic systems often based on the polymersase chain reaction (PCR) continues at NII and elsewhere. NII hopes that a monoclonal antibody against streptococcus-A will help to reduce the huge burden of rheumatic fever in India. Similarly the use of PCR techniques for identifying infection by M.leprae and M. tuberculosis will be of great value if they can be made usable in the field. Unambiguous ways of telling who is infected with what are a great impediment to the effective delivery of public health in India.

Temporarily, NII provides house-room for another laboratory active in many of the same fields — the New Delhi component of the UNIDO International Centre for Genetic Engineering and Biotechno-

logy (ICGEB), which has overlapping interests. But the intention is that this half-laboratory (the other half is at Trieste, Italy) should move into a new building nearby when that is completed next year.

For practical purposes, the Delhi component divides its effort between problems concerned with human health and those of plant breeding by genetic manipulation. One of its functions is to provide training courses for people from the developing world, but there is also an international staff.

Like others, the laboratory is excited by the recognition that the form of hepatitis known as 'non-A, non-B' is often, in south Asia, a distinctive microorganism now known as hepatitis E. This accounts for 40 per cent of the cases occuring in India, but already there is evidence that very different strains are prevalent on the two coasts of the Bay of Bengal.

In relation to hepatitis-B, the laboratory has already made some progress towards the cloning of a multi-segmented peptide consisting of several peptide segments recognised as antigenic epitopes. Now it plans to clone the whole genome of hepatitis-E so as to identify the antigenic elements. At the same time, it is working up diagnostic tools for distinguishing between the various infectious forms using PCR techniques.

Technically, the UNIDO laboratory is simply an international laboratory (or half of one) that happens to be based in India. In reality, with the coming and going of people from overseas, it is bound to be of great value to an enterprising research community that is both scattered within India and more isolated from the rest of the world (by diffidence as much as time zones and the cost of travel) than it cares to think.

The Indian community of those seeking to apply modern biology to India's public health problems is substantial and, now, robust. Apart from the government laboratories, many of the universities and indepedent research institutes have much to contribute. But the CSIR laboratory, the Centre for Cellullar and Molecular Biology (CCMB) at Hyderabad has a pivotal role, perhaps not yet explored.

The laboratory has a strange (and recent) history. The idea was that of Dr Pushpa Bhargava, originally a member of the regional laboratory at Hyderabad, now renamed the chemical technology laboratory. Over a period of ten years, Bhargava fought the bureaucracy for a general purpose laboratory encompassing molecular biology and biotechnology. But under his successor as director, Professor D. Balasubramanian, the laboratory has become a research laboratory as much concerned with basic as with applied research.

Between them, the members of the laboratory have interests in most of the quickly moving fields of molecular biology, from gene transcription to genetics. There is even a substantial theoretical interest. But the laboratory also reckons to produce a steady stream of practical benefits — improved fish feed and even genetically improved fish, for example, as well as the materials for an Indian DNA-fingerprinting system.

Even so, CCMB is less likely than other CSIR laboratories to be able to raise 50 per cent of its expenditure from outside sources by 1996, as Delhi has decreed. Indeed, Balasubramanian has difficulties even now in paying the laboratory's electricity bill. The question not yet answered is whether this laboratory should be recognized as a national source of modern expertise (and of trained people) for a still evolving national research network in molecular biology. Meanwhile, Bhargava sulks in his tent, embarrassingly nearby, saying that his dream has turned to dust and ashes.

The Institute of Microbial Technology (known as IMTECH) at Chandigarh will have none of these difficulties. One of the two institutions run by the Department of Biotechnology (the other is the NII), the laboratory began life in 1984 in an empty industrial building with staff recruited from the distinguished Drug Research Institute at Lucknow. Now it houses India's Microbial Type Culture collections, and is distinguished by the substantial pilot-plant fermenters it operates in the national interest.

One of the biotechnical ventures already in commercial use is a process for producing alcohol more efficiently from molasses, the standard raw material in India. The trick has been to breed a strain of yeast that is tolerant both to high sugar concentrations in the starting material and to high alcohol concentrations in the product. Although the savings consist largely in the cost of energy for operation fermentation process, they appear to be substantial enough to have persuaded alcohol manufacturers to take them seriously. □

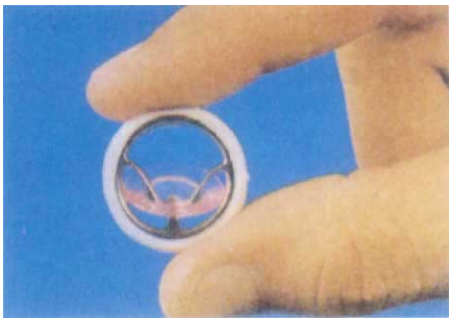
A valve that could do good

WILL India become the world's chief source of artificial valves for the human heart? There is at least a chance of that. Everybody who knows about this invention believes it will win an important place for itself in surgical practice. The National Research Development Corporation, which has acquired the rights in the invention from the Sree Chitra Tirunal Institute at Thiruvananthapuram, has already licensed TKK Pharma in Madras to make 10,000 valves a year. Enquiries about licences from outside India are, for the time being, being politely shelved.

N. K. Sharma, the corporation's managing director, is lyrical about the outlook. The essence of the advantage of the Indian valve is its cheapness. Valves imported into India cost Rs100,000, but the Indian equivalent costs only Rs2,500. The trick is the simplicity of the device, which consists of a specially manufactured disc of polyethylene oscillating in a circular aperture.

Materials scientist Dr S. Ramaseshan, previously director of the Indian Institute of Science at Bangalore, but now at the Raman Institute, is one of those consulted in the search for ways of ensuring that the polymer discs would be free of roughness even on a macromolecular scale, which would stimulate precipitation of macromolecules from the blood and perhaps even clotting. But Ramaseshan says that it has taken even greater ingenuity to enclose these discs in their titanium cages.

The immediate plan is to produce 10,000 valves a year at Madras, testing one in ten in dogs to establish reliability.



Promising much.

Sharma points out that even that production rate is a small proportion of the number of people in India with heart valves damaged by rheumatic fever, which affects an estimated 100,000 children a year.

In a sense, the marketing of these devices will be a crucial test of India's claim on technological prowess in the modern world. So far, its importance seems to have been fully appreciated. Sharma's people have given themselves the coming year to prove to themselves that their manufacturing methods do indeed produce reliable valves, and whether themselves to apply for regulatory approval, or alternatively to let licensees take on that chore.

It is generally agreed that a reputation for unreliability in manufacturing must either be avoided (by licensing a good idea) or met head-on, by indigenous manufacturing of reliable products. If all goes well, the Thiruvananthapuram heart valve could well make India's NRDC as prosperous in the 1990s as the cephalosporins made its British namesake in the 1960s. □

Telecoms and watches sustain chips

THE most prominent feature of the Semiconductor Complex at Chandigarh is a reflecting pool the size of a football pitch surrounded by splashing fountains. On a morning last month when the habitually clear sunlight was reflected from the pool, Dr M. J. Zarabi, the chairman and managing director of the company that owns the plant, was quick to say that the reflecting pool is functional as well as decorative; it is meant to cool the water used on the silicon production lines.

But they are not as active as they should by now have been. Semiconductor Complex Ltd owes its existence (since 1983) to a decade's earlier brooding by the government of India about India's place in chip-making technology. Despite earlier interest by the Atomic Energy Commission in silicon devices (some of it in collaboration with the Dutch company Philips), a committee first formed in 1971 eventually recommended that India should build its own capacity for designing and manufacturing silicon chips.

Work began on the Chandigarh site in 1983, with technical assistance from American Microsystems Inc. (AMI) with the technology of 5-μm devices — and then the vapour deposition plant caught fire and burned down at the end of 1987. (There was an electrical short-circuit, but nobody was injured in the fire).

That has been a huge setback for the company. The previous government agonized over the proposal that the damaged plant should be rebuilt. Meanwhile, the Semiconductor Complex has arranged to fabricate silicon wafers at AMI's Austrian plant, using its own workers relocated for the purpose to gain experience. But in 1991, with the accession of the Rao government, the decision to rebuild was taken. Now the crucial fabrication plant is being restored again; predictably, when it is commissioned a year from now, it will be a better plant — and will have cost the equivalent of $70 million.

But Zarabi is not depressed. He has worked at the plant since the beginning, but only recently as the boss. Meanwhile, he is proud of two things. First, the company has been making a profit on its operations since 1989 (before allowing for capital charges), largely on the strength of silicon chips designed for telecommunications switches (made elsewhere in India) and the timing mechanisms of watches, which are sold to watch manufacturers using different labels. Annual sales now amount to Rs670 million.

Zarabi's second source of pride is that the Semiconductor Complex, building on the 5-μm technology acquired from AMI, is now able to manufacture devices with a resolution of 1.2 μm. The hope is that the company will be able to make chips with 1.2-μm resolution by the middle of this decade.

Against the time when the burned building is restored, the company plans a number of distinctive products. For one thing, it is busily working on the design of chips that will operate at d. c. voltages higher than the usual 6 or 9 volts. Its concession to the decision in the current five-year plan that potable water must be a national priority is to devise a biosensor card to monitor the purity of water supplies.

India's telecommunications technology could fit the bill for export markets

The constitution of the Chandigarh company is typical of the standard Indian pattern of nationalization. Formally, Semiconductor Complex Ltd is a public company, but its shares are held exclusively by the government or by other nationalized companies. Similarly, the directors of the company are appointed by the Department of Electronics in Delhi. It is one of several companies similarly constituted and all active in the field of silicon electronics.

Zarabi nevertheless is unshaken in his belief that indigenous chip-manufacture is essential for India's well-being. He acknowledges that the acquisition of 5-μm technology was essential to get started, but says he would have resisted any suggestion that the finer chip constructions should not have been developed at Chandigarh. But "who are we in this part of the world to say we shall never need other people's technology?" Zarabi's position is that India will always, in technology, follow a mixed policy of "development and buying"; but he insists that it is possible to buy from a position of strength if there has been indigenous development as well.

Much the same is said at the Centre for the Development of Telematics (C-DOT) at Bangalore, another nationalized venture of the Indian government for the development of digital telephone switches which is also exactly ten years old, which is primarily a designer of equipment manufactured by others. More than 30 Indian companies have been licensed to manufacture different products, some of them with no previous experience in the

Bangalore central to electronics

BANGALORE has become the centre of India's electronics industry not simply because it boasts of the Indian Institute of Science, but for a more mundane reason: there seems to be less dust in the atmosphere than anywhere else in India.

As well as C-DOT and a growing number of small private companies, the city also houses one of the largest electronics manufacturing plants in India, Bharat Electronics Ltd, a government-owned defence contractor now seeking a place in the civilian market.

The company's claim that its military communications equipment is among the best in the world is well attested. Ten years ago, this seemed to offer a basis on which India might enter the international consumer electronics market, which never happened.

The plant is everything one would expect of such a place. By Indian standards, every room is a clean room, but most are technically so, with shoe-changing entryways and every employee in a white hat and coat. (Bharat has even devised a uniform into which everybody changes, which, for the women employees, consists of a fawn-coloured sari.) At Bangalore, the company also follows enlightened employment practices: there is a crèche for more than 100 toddlers accompanying their mothers to work, as well as a fleet of buses to take everybody home at the end of the day.

With 20,000 industrial employees scattered over 10 sites in India, Bharat is an important part of the Indian economy. But why so scattered? That may be one of the hidden costs of being a government-owned company. Its latest plant, in the foothills of the Himalayas, has been put there to generate employment. □

telecommunications. Annual sales of C-DOT equipment are said now to amount to R7 billion a year, on which the company earns royalty income of Rs200 million.

Again, the organization seems to have begun with a licensing deal with Siemens on software for digital switching gear, but it now reckons to be technically independent (thanks to several hundred software engineers at its New Delhi office). The chief user of its products is the Indian telecommunications system, which now has 7 million subscriber lines installed but which is growing at a remarkable 25 per cent a year.

C-DOT's products inevitably reflect the diversity of India's telecommunications needs. The larger switches are meant to handle 10,000 lines or more, but the company also worries about the villages, with the result that some switches are meant to distribute digital signals to single village telephones by means of high-frequency radio carriers. It is striking that the switches, whether small or large, use similar design tricks to provide redundancy, for example.

Technical skill abounds at the Bangalore development shop. The line managers are confident and as bright as paint. And, being a nationalized company, its directors say they are not dismayed at losing a substantial number of their engineers every year. Is that not part of the process of deepening India's technological base?

So has the past decade's investment been worthwhile? C-DOT has no doubt that, without an indigenous capacity to manufacture digital switches, India would be held to ransom by overseas manufacturers. But it is confident enough of its own products to look forward to buoyant export sales: now we're competitive, especially on price. So C-DOT is talking to the telecommunications authorities in other developing countries, hoping for great things in countries such as Indonesia and others in Africa, which have similar needs to those of India.

How good is that calculation? Whatever India's distrust of the market, the market will be the final arbiter of success. On the face of things, C-DOT has a technology that others will find attractive. It has not yet come up against the questions India habitually asks of manufacturers in the West, that of whether they will license local manufacture of equipment. It will be interesting to see whether India, with its long-term worry about the need of jobs at home, will be as compliant on this point as it expects others to be.

But is there not always the domestic market? That is where even the first stages of economic reform may hurt, for they allow overseas manufacturers to manufacture their wares in India. But local people are convinced that overseas companies will not be able to hire Indian engineers for salaries as low as the Indian companies offer.

But even that supposes that the technology of telecommnications will remain what it now is. At the beginning of December, a delegation from Motorola was due in India to seek membership in a satellite-based global cellular telephony system. Given the comparatively low penetration of telephones in India and the problem of the villages, might it not be simpler to start all over again with a cellular system. The offer will probably be turned down on this occasion (they only want our money). But it will not be the last opportunity of its kind. □

Data delivery to the people

How many hand pumps are in working condition in Kuppam village in Tamilandu state? How many infants died of cholera in Kavaratti (an Indian island in the Arabian Sea) in May 1993? Five years ago, central government officials in New Delhi would not have dared to ask these questions. The information either did not exist, or it travelled too slowly to be of any use.

Not any more. The task of governing a country as large as India has become easier thanks to the satellite-based computercommunication network being operated by the National Informatics Centre (NIC) in New Delhi. Now, officials can key in their queries on NIC terminals on their desks. NICNET, as the network is called, brings to their screens answers to such questions as the number of cows in a particular village or the length of roads in need of paving.

Information on vital sectors of the economy such as health, agriculture and water are now stored inside NICNET's super- and super-minicomputers distributed across India. The database is constantly updated. Planners and decision makers in state capitals and central government can access these databases using home-made antennas looking at India's INSAT-ID satellite.

According to N. Seshagiri, director-general of NIC and architect of the concept, "NICNET is the largest data network of its kind used for government administration". It now links central government ministries in New Delhi with 32 state capitals, and the state governments with all the 500-odd districts — the basic administrative units of the country.

In the next phase, the network will be expanded to the block level. India is divided into about 6,000 blocks each consisting of 115 villages in a 12-km radius. In 1996, NICNET will have in its computers basic data pertaining to each of its 600,000 or so villages.

Seshagiri says that NICNET, which provides electronic mail, file transfer and bulletin board services, has helped to bridge the communication gap between centre and state governments, and between district administration and the state capital. "Every day 240,000 transactions take place", he says. From February 1994, officials will be able to hold video conferences, a saving on travel expenses.

The government has invested about £50 million in NICNET, but the unquantifiable returns are proving much greater. For example, officials in Delhi are using the network to monitor the progress of centrally funded development projects worth £30,000 million. State governments are also finding NICNET extremely useful. For instance, a Karnataka agricultural bank recently discovered that by using NICNET it could process the monthly statement from all its 177 branches in the state in one hour on the first day of the month. Previously, the task had taken 20 days for 12 people working 12 hours a day.

NICNET came in handy in 1992 for reporting results of the general elections and during the 1991 census. For the first time, the census data pertaining to 850 million people were processed and brought out in report form in less than a year. And after this year's devastating earthquake in Maharashtra, NIC's mobile Earth station established 24-hour link between rescue teams and the state capital.

Although only government departments can now use NICNET, Seshagiri is soon hoping to get approval for opening it to nongovernmental agencies and foreigners. Though its services to the ministry are free, NIC earns £2 million annually from public sector companies and is hoping to earn more by selling specialized databases.

As things are, the general public can use NICNET's special terminals, installed in several cities, to obtain such information as hotel and medical facilities or railway timetables. Another database accessible to the public is about proceedings in the supreme court and high courts. By paying £1, a litigant can use an NICNET terminal anywhere in India to know about the status of his case and when it is posted for hearing.

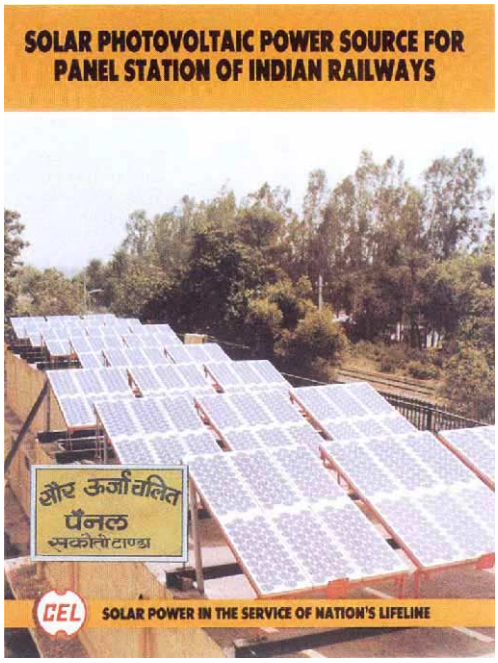
According to Seshagiri, with the evolution of NICNET, much of the information previously accessible only to a few government servants has come into the open. This is a good thing he says "because transparency of information will lead to a situation where officials cannot afford to be corrupt".

K. S. J.

An energetic company

ELECTRONIC devices, without moving parts, are unspectacular to look at, but Central Electronics Ltd, at Sahibabad in the eastern suburbs of Delhi, has made electronics fun. Its stock in trade is the production of solar cells for the production of electricity. Its director, M. R. Narayanan, is a retired Indian Amry brigadier with interests on the signals side who seems to have been captivated by turning India's endless sunshine into power.

That the trick is possible is plain for all visitors to see. In a crudely constructed shed on CEL's substantial campus, there is a display of working models, many of them aimed at India's village folk, all driven from a solar collector on the roof. For example, there is a kind of portable hurricane lamp whose battery is recharged by the solar collector when not being moved.



Few applications escape CEL.

And, inevitably, there is a solar-powered television set, which in real life might be the medium for disseminating improving information to the villages. CEL has also discovered the need for rudimentary lighting systems, perhaps in evening classes, and has built a room to show off that prospect. For semi-urban India, CEL's brochures also show a solar-powered petrol pump.

But the neatest device, entirely consonant with Narayanan's first career, is a portable solar collector consisting of half a dozen square-foot silicon surfaces that fold neatly into a canvas square that can be fixed to a backpack. The idea is that, when laid out in sunlight, they produce enough power to charge an accumulator in a reasonably short time, thus making it possible to operate a radio from some remote location. The benefit, for soldiers, is that a spare accumulator does not have to be taken along.

That is CEL's shop window. Its strength is its experience in hitching up solar collectors to the devices that will use the power they produce. Working on the principle that there is plenty of sunlight anyway, the laboratory is aiming at achieving reliability rather than striving for the ultimate in efficiency. (It reckons that its silicon modules can convert 13 per cent of the sunlight they collect into electricity.) At one point during a walk through the production plant, Narayanan lifted one of several modules of silicon cells from a stack and dropped it face-down on the concrete floor just to demonstrate its ruggedness.

In practice, it appears that the basic (but still evolving) design of the standard crystalline silicon module (which is a little larger than a small floppy disk) is less important than the arrangement of the small components into a power-producing module (the spaces in between, and their texture, matter) and the collection of the current. CEL has for the time being settled on a modular solar collector just over one square metre in area, which can produce 3.3 kW of power at its peak. That is enough to pump water from a shallow depth for about 2 hours a day.

The intricacies, which are not part of high technology, come with the processing of the power. To iron out night and day (and the rare occasions when the Sun is not shining) power must be stored in batteries and even converted from d. c. to a. c. The accumulator is a bulky (and costly) item, and may not last for long if it is overcharged or allowed to run too flat, as is the convertor.

So how quickly will Narayanan's solar cells help to defeat the threats of acid rain and global warming, on which he has set his sights? The soldier turned salesman is ever ingenious in the search for customers; at least one off-shore oil-drilling rig has been fitted out, but there is also, alongside the campus helicopter pad, the essence of a 1,000 kW installation that may just about supply CEL with power. There is also a plan for 10 MW station in an isolated township nearer the Himalayas.

But the Eighth Plan is cool on the prospects for solar voltaics. Cost, it says, continue to be a major constraint in the expansion of the programme. In a familiar note, the plan says that energy conservation is still the more economic course to follow. □

A future in bamboo

MAKING bamboo plants flower at will may seem an esoteric enterprise, but it has become one of the chief interests of the National Chemical Laboratory (NCL) at Pune. The point is that, in the wild, bamboos flower only at unpredictable intervals, perhaps after 30 years or more. The result is that conventional plant-breeders are hampered, to say the least of it, in seeking crosses between different varieteies and species.

But no longer. Since 1990 and the publication of a paper (*Nature* **344**, 355; 1990) by R. S. Nadgauda, V. A. Parasharami and A. F. Mascarenhas showing that the flowering of bamboo could be induced at will, NCL has been up to its eyes in the exploitation of the phenomenon. The objective is to use the system for interbreeding between varieties and species using pollen and seeds collected from flowers induced to flower artificially.

The technique, which relies on an empirically derived broth whose active principles have not yet been characterized, has been made to work with most (but not all) bamboo species, from India and eleswhere. The group at Pune has shown that immature seeds collected from induced flowering can be fertilized and mature plants grown from resulting embryos. But attempts to induce flowering of mature plants have so far failed, suggesting that the technique will work only *in vitro*.

That should suffice. Bamboo is a versatile crop widely used in India for construction and as fodder. It should not be long before the NCL has improved specialized strains on the market, when it will be interesting to see whether it follows Western seed distributors in seeking royalties on the crops that arise.

But Dr A. R. Mashelkar, director of the NCL, has other fish to fry – one of them a scheme to manufacture carbamate pesticides in India by processes other than that responsible for the tragedy at the Union Carbide plant at Bhopal exactly ten years ago. (The process is now banned in India.) The alternative involves the condensation of alkyl carbamates with phenols, but Mashelkar has now put his faith in still another alternative, involving the reaction of high-pressure carbon dioxide with amines. Green chemistry is his slogan.

Equipped with large-scale pilot plants, NCL works closely with the chemical industry in India on the development of novel processes, many using zeolite catalysts. Mashelkar has much to say about globalization. He insists that India is capable of meeting the challenge in chemical manufacture; The mood is there, the question is how fast you can go. □

Space — more a question of pride

DOES India need a space programme? Ask Dr U. R. Rao, the chairman of the Space Commission, and you are left in no doubt: India needs a capability to launch satellites just as much as it needs, say, potable drinking water in every village. Indeed Rao, by his own voluntary admission a "powerful" man, goes a long way to suggest that there will not be drinking water in every village without a space programme.

And an independent one. Rao, who is charming and intelligent as well as powerful, does not want India to be pushed about. He offers as illustration of how India can be ill-treated the case if the recent Russian refusal to transfer the technology for building cryogenic (liquid hydrogen) rocket engines, widely believed to be the result of US pressure. "Who in his senses would think that people would use cryogenic engines in military rockets?", he asks.

Speaking at the satellite laboratory of the Indian Space Research Organization (ISRO) at Bangalore at the end of last month, Rao was clearly looking forward to the visit the following week of a delegation from Russia bent on renegotiating the contract to transfer the cryogenic technology. Contrary to the expectations of his colleagues that he would make a fuss, he declared ominously that he would "listen to what they have to say". Meanwhile, to judge from the public construction tenders being advertised in the newspapers, India is pushing ahead with its cryogenic facility, with or without Russian help.

So is there a military rocket programme? Rao reports a tense conversation with US defence people complaining that "half a dozen" ISRO engineers had recently joined the defence ministry. "I told them that many more had gone to work for NASA in the same period." Even so, it is only prudent to suppose that there are military as well as civilian rockets, and that the knowhow gained in developing the latter is somehow transferred. But ISRO's activities are otherwise transparent. And Rao has a persuasive case for linking Earth satellites with water for the villages.

India's largest launcher, for 1,000 kg payload, launched in September.

One of the extensive ISRO satellite family

India's practical interest in Earth satellites is two-fold: communications and Earth observation. There are ample practical successes to boast of, not least the use of a geosynchronous transponder to provide warning of impending typhoons to a clutch of unattended monitoring stations on the coast of the Bay of Bengal. Rao says that the Disaster Warning System saved 20,000 lives from a typhoon that struck in November 1991. (This is the typhoon season of the year.)

Ingenuity, with the sense of improvisation that inevitably brings, has been a hallmark of the Indian space programme ever since Vikram Sarrabai's people at Ahmedebad in the late 1960s temporarily used a US satellite to send television signals to a few Indian villages. Rao is now hoping to institutionalise that old "SITE" project with six to eight high-power C-band transponders in a satellite system called GRAMSAT – "Gram" is Hindu for "village" – meant to get rid of illiteracy and provide continuing education as well as open university instruction, in a dozen or so native languages. The hope is to reach 90 per cent of the population through 80-cm dishes within five or six years.

ISRO's Bangalore laboratory has become the home of many other semi-autonomous providers of satellite-based services. In a country so big, mere geography is best done from Earth orbit, which the schemes for monitoring land use (by different crops, for example) and flooding (during monsoons), and for their improvement and abatement respectively. ISRO cut its teeth on US and then French ("SPOT") satellite images, but now proudly collects its own from the IRS-A (launched by an Ariane rocket in 1991) and IRS-B satellites, each of which collects images in one infrared and three optical wavebands.

The neatest application so far is the use of sea-surface images to spot the dividing lines between warm and cold offshore water, where fish are most likely to be found. A warning system has been set up, which has in turn provoked a typically Indian quarrel; the smaller fishermen are asking that they should be given the first advice because their boats are smaller, and take longer to reach the designated fishing opportunities. That is the kind of imagined injustice that the politicians will be eager to get their teeth into; ISRO, which wants to resist, is lucky to have Rao on its side.

The launcher business is India's real worry. Its ambitions have always lagged behind its hardware. At the outset of the space programme in 1963, it borrowed sounding rockets and learned the hard way about telemetry, but was building its own by the end of the decade. The decision to build a four-stage rocket (SLV-3) capable of launching 40-kg satellites was taken in 1969 and carried through by 1980 — by which time it was clear that 40 kg in near-Earth was to little to serve much purpose and a decision had already been taken to build and even larger four-stage rocket — ASLV (for Augmented Satellite Launch Vehicle). No there is an even larger one nearing completion, called PSLV (for Polar Satellite Launch Vehicle); there was an unsuccessful first launch earlier this year. But even that will not put substantial communications satellites into geosynchronous orbit, whence the Russian cryogenic fourth stage.

For a country established as the firework Eldorado of the Third World – Indian fireworks sell even in China, but are believed to owe their cheapness to the low wages of the children used to make them around southern cities – that is galling. Rao says it is even worse: now that the Russians have agreed not to undercut the United States in launch costs, Ariane's costs have also come into line. He is also peppery about the asymmetry of the contracts the commercial launchers demand: customers incur penalties if their satellites are delivered late, but may then have to "twiddle their thumbs" for months while the launching rocket is made ready.

Given all that, the chances are high that India will continue to develop its own launch vehicles, to the irritation of developed countries who ask that the money should be given to the poor instead. That opinion neglects India's jealous pride in its independence, which in this context seems appropriate. The hope, instead, should be that the price will not be excessive. □

GATT adds fuel to patents controversy

INDIA remains as suspicious as ever of suggestions that it should follow procedures on the protection of intellectual property, patents and the like, advocated by the industrialized West and, in particular, embodied in the latest version of the General Agreement on Tariffs and Trade (GATT) now being negotiated.

One of the contentious issues is India's patent legislation, which exempts from local protection products (but not processes) judged to be of particular value to the health and education of people in developing countries. But India is particularly disturbed that the new GATT regime (of which it wishes to be a part) will require protection for strains of plants developed abroad in ways that make no sense to Indian farmers.

So farmers and agricultural scientists in India are upset over the government's decision to sign the GATT treaty. They are not convinced by official explanations that signing the treaty will not harm their interests.

The draft treaty proposes the patenting of microorganisms and also stipulates that signatories should provide "for the protection of plant varieties either by patents or by an effective *sui generis* system or by any combination of these". It is this part that farmers fear will affect their traditional rights to save seeds from one crop to the next.

While farmers in the industrialized West are habitually dependent on private companies for seeds, 80 per cent of seed supply in India is within farmers' control. They reproduce, exchange, breed and sell their seed and plant varieties. Now they fear that they will end up buying seeds from multinational corporations holding the patents and protective certificates on plant varieties and seeds. Agricultural researchers are also concerned that they will be required to pay royalties on protected varieties they may wish to use for genetic enhancement, even by traditional methods.

"We shall have riots on our hands if we ask the farmers not to keep their seeds from one crop to the next", one politician said last week. "Is that what the West really wants?"

The merging of farmers' and scientists' interests has created a strong anti-GATT movement. Cargill Seeds India Limited, a subsidiary of the US company of the same name, was twice physically attacked last year in what was said to be a warning to multinationals to keep out of the seed business in India. With the impending signature of the GATT treaty, the movement last week announced plans for intensifying its agitation. The new protest includes a boycott of seeds produced by multinational companies and the encouragement of seed exchange through farmer-run seed banks throughout India.

But the Indian government says that farmers and scientists are over-reacting. In a statement in parliament last week, commerce minister Pranab Mukherjee said: "There is no obligation on us to patent seeds and we do not intend to do it .... but some sort of protection for plant breeders is in our own interests." In the same breath he announced that the government will introduce legislation incorporating "the farmer's right to retain his seed from one crop to another and to exchange seed in traditional manner".

But Vandana Shiva, a United Nations consultant and expert on these issues, says there is no way the government can ensure farmers' rights and researchers' privilege without actually scrapping the offending clause in the treaty. She is angry because Indian negotiators at the Uruguay round did not press for such an amendment.

The *sui generis* system India intends to adopt is that codified by the 1978 convention of the Union pour le Protection des Obtentions Végétals, or UPOV. This does protect farmers' rights to save seeds for their own use, but it is restrictuve in the sense that breeders starting with protected varieties cannot commercialize the results of their research. "In simple terms," said a senior agricultural scientist, "it means that India cannot make another green revolution without paying enormous royalties to owners of patented plant varieties.

It seems especially ironical that the proposed regulations should apply to proprietary interests owned by corporations overseas in means of using varieties of plants native to India itself. The controversy is amply illustrated by the case of the neem tree, native to India and traditionally used for a variety of purposes.

The neem has been processed in India for centuries by techniques involving grinding, heating with a variety of solvents, refining, extracting and diluting. Among other things, these processes have been used commercially to yield neem-based pesticides.

But "the West pirated this knowledge of the Indian community," says Vandana Shiva, director of the Foundation for Science, Technology and Natural Resource Policy in Dehra Dun, who complains that broad-based patents held by W. R. Grace, a Florida-based company, mean that Grace now holds the rights to grinding, heating, refining, extracting and diluting parts of the neem to produce pesticides containing azadirachtin.

Under the GATT, the Indian community cannot charge Grace with piracy, says Shiva, because the rights of the community to intellectual property are not recognized. But Grace will have the right to charge India with piracy when Indians, under the "reversal-of-burden-of-proof" clause, would have to prove in court that they have not copied Grace's process. The people of India will be treated as guilty of misusing Grace's patents until they prove themselves innocent.

Whether or not the traditional uses of the neem tree are ever brought to court, Indians are alarmed at the far-reaching implications of certain clauses of the Trade Related Intellectual Property System (TRIPs) proposed by the GATT treaty. They are particularly incensed at the effect of the clause making it incumbent on the defendants in a protection suit to prove their innocence.

In the case of the neem tree, if India were to fail to recognize Grace's patents, the United States would be able in principle to levy tariffs on imports from India until the national economy were sufficiently hurt to force recognition.

K.S.J.

India pushed about?

Rights to intellectual property is a contentious topic. Many in India passionately regard the issue as a means of coercion by 'the West' comparable with the diplomatic pressure brought to bear on India by its leadership of the 'non-aligned' group of governments.

But it is a more complicated issue that many will allow. Thus Indian patent law does not allow for the patenting of products, but only of the processes for manufacturing them, while the duration of patent protection (14 years) is less than elswhere.

Moreover, the patenting of naturally occurring materials is not allowed, casting doubt on attempts being made elsewhere to patent human genes or parts thereof.

On seeds, the present practice in India is that the agriculture ministry supports the Indian Institute for Agricultural Research as well as the network of research stations and extension services required to develop improved seeds and to see that they are properly used. No attempt is made to recover from the eventual beneficiaries the costs of research and development, although in principle the government could attempt to do this by adjusting the prices it will pay for crops that are surplus to farmers' own requirements.

Branching out in chemical systems

THE most improved laboratory in India must be the Indian Institute for Chemical Technology (IICT) at Hyderabad. Ten years ago, as the Hyderabad Regional Laboratory, it was one of the worst. The difference is not so much the change of name, but the arrival as director of Dr A. V. Rama Rao, a vigorous no-nonsense organic chemist of distinction.

But not all traces of the old laboratory have vanished. Indeed, the elaborate pilot plant then constructed to develop a process for the gasification of Hyderabad's peculiarly high-ash coal is still in place, but a little rusted. Rama Rao says that on his arrival in 1985, he insisted that the future of the project should be decided quickly, one way or the other. It took two years to show that the process would not be economic.

Meanwhile, IICT has turned its attention to more tangible projects, one of them a different process for the manufacture of AZT, the drug widely used in the treatment of AIDS. The inspiration of that project was the calculation that the Wellcome Foundation's raw material are *D*-ribose (which costs $200 a kg) and a-thymidine ($800 a kg), but that *D*-xylose, at $20 a kg, could be a source of the thymidine.

IICT has licensed the Indian manufacturer CIPLA to make the resulting AZT in India. Arrangements to manufacture the drug in Thailand and Brazil have also been agreed, although the arrangements stipulate that CIPLA should be free to export to both countries if it can do so more cheaply than the local manufacturers.

Otherwise, IICT is busy with the development of processes for the manufacture of the approved alternatives to the chlorofluorocarbons (CFCs) used as refrigerants and banned by the Montreal Protocol of the ozone-layer convention. Rama Rao says that having developed an economical process for producing the hydrogenated CFCs, he is now locked in argument with the World Bank over the capital cost of a production plant, which he estimates at Rs50 million.

He is annoyed that the fund established to compensate developing countries for their use of alternatives to CFCs, which is administered by the bank, will meet the extra costs of use, but not the capital costs of making them, which would be cheaper in the long run. Gently, he suggests that this is a way of making India a perpetual customer of producers in the West.

For the rest, IICT has a host of other projects to launch onto the market. Rama Rao's own research interest in antitumour agents has been led to processes for the manufacture of materials such as vinblastine and vincristine. Half of India's annual production (50 tonnes) of vitamin B6 by an IICT process is exported. Indeed, Rama Rao claims that the annual value of the drugs being manufactured by processes developed by him and his group amounts to Rs2 billion (say US$65 million). □

Thirty-dish array aiming to see further than VLA

ON the western edge of the Rajasthan Desert, on a site 80 km north of Pune, Indian radioastronomers are building a radiotelescope that will be the most powerful of its kind in the world. If the current pace of construction can be kept up, the metre-wavelength instrument (GMRT) will be ready late next year or early 1995. Will that give India's small but renowned group of radioastronomers a decisive advantage?

Everything will depend on how well the instrument works. But the project has excellent credentials. It is being carried through by the National Centre for Radio-Astrophysics based at the Tata Institute for Fundamental Research in Bombay, and is the brainchild of radioastronomer Govind Swarup, the director of the national centre. The site has been chosen for its freedom from man-made radio noise, which is especially troublesome at comparatively long wavelengths.

In many ways, GMRT compares with the US radiotelescope called the Very Large Array (VLA) in New Mexico. When completed, it will consist of 30 fully steerable parabolic dishes of 45-m diameter spread out along the three arms of a Y-shaped configuration. Four dishes are already in place and a new one is added every two weeks. In terms of resolving power, the array is equivalent to a single gigantic dish 25 km in diameter, Swarup says. The chief difference between GMRT and the VLA is that the latter's dishes are movable on railway tracks as well as steerable.

The decision to build a telescope designed for metre wavelengths is a mark of India's confidence in radioastronomy. One objective is to look for the nebulae in the early Universe from which it is supposed that galaxies were formed. Such objects, consisting predominantly of neutral hydrogen, would have emitted the familar 21-cm wavelength radiation characteristic of hydrogen clouds in nearby galaxies. The expectation is that this radiation will be red-shifted by the expansion of the Universe so that it has an apparent wavelength between 1 and 2 metres, or within the GMRT bandwidth of 30 to 1,500 MHz.

One of the planned 30 dishes.

But Swarup also reckons that the new telescope will be used for searching the radio-sky for pulsating radio stars whose repetitive signals are even more frequent than the recently discovered millisecond pulsars.

The design of the new instrument is distinctively Indian. To save on materials cost, for example, the dishes are made of see-through wire mesh instead of solid metal. The total cost of the entire project is US$15 million and, except for the German bearings, has been entirely fabricated in India. Overseas astronomers will be offered 10 per cent of the observing time.

That echoes Swarup's original hope that the new telescope could have been built as an international project on an equatorial site (not in India) so as to cover both hemispheres of the sky. When functioning, it will more than keep India's radioastronomers busy with observations. (There is also a substantial radiotelescope at Ooty, in South India, and a millimetre-wave telescope at Bangalore.) The project's well-wishers outside India marvel at two of its outstanding features: the small construction cost and the belief that it will be possible to weld such a complicated array of radio-dishes into a finely tuned observational instrument within the next twelve months or so. □

Science and the developing world

Science and technology have a recently developed affinity with the development of the poor countries of the world, but have few effective ways of winning benefit for either partner in the relationship.

WHY is science now so much concerned with development to improve the lot of the poor countries of the world? It has not always been so. The people who followed the first explorers of the globe were mostly traders, more concerned with the artefacts and natural products that could be gathered than with those who lived in far-off places. They were followed by organizations such as the British East India Company, created to safeguard earlier investments, which found themselves inadequate for their designated tasks without the military backing of their governments. Missionaries followed in the nineteenth century. Except for South America and Central Asia, most of the developing world has been through that cycle.

What follows is an extended postscript to the preceding survey science in India, based on the umpteenth visit in a quarter of a century, but its purpose is different. The state of science in India can be described, but what can be learned about science from its long-standing devotion to the cause of development, however fitful and wrong-headed it may be?

For what it is worth, the first steps towards the exploitation of science in India were taken by colonial administrators, not scientists. The military, when they came to occupy the country, needed reliable maps and recruited surveyors to do the work. But telling the longitude was a matter for astronomers, who knew how to operate transit instruments. That is the origin of the first observatory in India, at Madras.

The next steps were also taken by colonial administrators, but largely in this century, copying in India (and other colonial territories) research organizations set up at home to track the weather or to win better crops. Except in what is now the United States, the colonial administrators seem never to have supposed that intelligence might make the occupied territory a better place than were the place they had come from.

Even in retrospect, it is remarkable how recent is the concern of Western science with the ex-colonial territories. That is a phenomenon of this century, and almost of the past half of it alone. Of course, the League of Nations (a creation of the early 1920s) helped to put developing countries before the eyes of the richer world. Since the Second World War, the United Nations has institutionalized the cause of development, especially through its specialized agencies.

We forget how much of that has come about because of the positions taken by individuals. There is a host of Scandinavians who share the credit. Mostly, they were driven by political considerations, perhaps even by the argument that small countries like their own might be accorded more respect if countries even less fortunate were counted worthy of respectful (and not patronizing) attention. Science was not a vocal element in this movement until the Second World War.

Why have things changed? There are several components of an explanation, of which the chief is the intolerance of impotence in the face of soluble problems that understanding breeds. If a person (or his cattle) have no water, why not make a pump, drill a hole, and get some? If they have no food, then create an agricultural research council. And if a person's children are repeatedly dying of an infection, why not identify the organism and devise a prophylactic? We underestimate the importance of that tendency, patronizing (and out of tune with recipients' felt needs) though it may be.

Second, but much less important in science, is the strategic argument that an unequal world is an unstable world, subject either to open conflict or to equally debilitating demands on resources. Those who have read Gibbon's *Decline and fall of the Roman Empire*, claim that its message is not that the northern European tribes were seeking to destroy Rome, but that they wished to share in its prosperity. In present circumstances, of course, that piece of history could not repeat itself, although the emergence of China as the successor of the Soviet Union as the nearest entity to a second superpower, its low wealth per head notwithstanding, is a surprise.

The third reason for the interest of science in development is biological, but undetermined. Many years ago the English Catholic saint, St Augustine, faced with the difficulty of explaining why, if life on Earth is merely a preparation for a better life in Heaven, death should not be welcome, hit on the notion that people now alive have a regard for their successors. The modern statement of this position, as in Dawkins' *The Selfish Gene*, is that people have a vivid and instinctive regard for their fellows. At last month's Indira Ghandi symposium in Delhi, participants repeatedly announced that the cause of development is a "moral" cause. If they were referring to the principle of the selfish gene, well and good.

India is not a developing country in the strict sense, but two countries in the same territory, the poorer at least twice as numerous as the richer. It is amazing and moving to see the commitment of scientists from the richer part to the improvement of the other.

Yash Pal, a pipe-smoking physicist from Surrabai's northern stronghold at Ahmedebad, has spent a lifetime seeking to bring television (and now instruction) to the masses. C.N.R. Rao, the director of the Indian Institute of Science at Bangalore, is the chairman of a literacy organization with 100,000 volunteer teachers — and a budget equivalent to £4,000 a year. M.G.K. Menon, science advisor to several prime ministers of India, but now marooned as a member of India's Upper House of Parliament, says passionately that development for the poor part of India is at the top of his agenda. And India is awash with people whose spare time goes on rural science or simple instruction.

It is no reflection on the motives of these people, all of them friends, to say that the benefits they bring to the poor part of India also bring them a sense of achievement not easily distinguished from pleasure. Who says that good works should make the doers miserable? But it is also impressive that these *in situ* developers sense instinctively what developers from outside more often have to learn by instruction or even with greater pain, that the other world's solutions to poor people's problems are not solutions if the poor people do not embrace them. Villagers are also deserving of respect.

Delhi's administrators are usually more impatient. Believing as they do that India's development is a matter of putting technology to work, they repeatedly draw attention to the importance of "technology transfer", as if to imply that India will be unable to deal with its development until Cray Research Inc. has sent a set of blueprints for its latest supercomputer and trained a workforce to replicate it. India and the rest of the world deserves a more liberal deal on intellectual property, while advanced technology is not irrelevant to the condition of the poor, but that is to ask too much.

The better strategy, for those for whom development is an urgent need, would be to exploit the willingness of the technical community elsewhere to assist. Already, there is a substantial flux of technical people to training positions in the rich world; should not this be even greater? There is also a good supply, even an over-supply, of experts of various kinds, but the urgent need is of the simpler skills of teachers and even plumbers, or of the cash with which to employ them locally. India is lucky in knowing what it needs. How soon will other poor countries understand?

John Maddox

GTPases

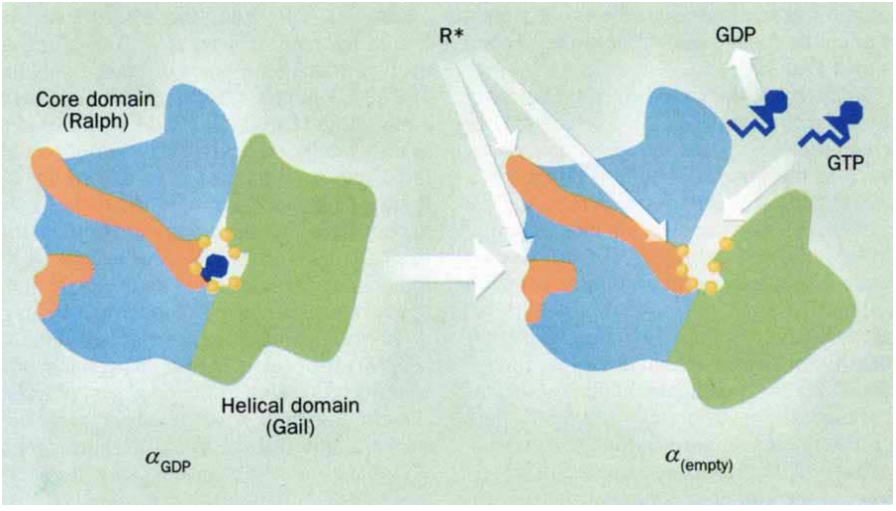
A turn-on and a surprise

Henry R. Bourne

WHY do biologists flock to gape at a new protein structure? For its usefulness and beauty, certainly — but also to satisfy a craving for surprise. The crystal structure of the α-subunit ($G_{t\alpha}$) of the G protein heterotrimer, transducin, described by Noel, Hamm and Sigler on page 654 of this issue[1], furnishes a beautifully satisfying and useful surprise. $G_{t\alpha}$ mediates vision in the retina by transmitting light signals detected by rhodopsin, the photon receptor, to cyclic GMP phosphodiesterase, the effector. The three-dimensional $G_{t\alpha}$ structure reveals the detailed anatomy of a G protein α-subunit, one of a family of specialized GTPase-driven switches that relay and amplify signals from cell-surface receptors to effector enzymes and ion channels. A striking structural difference between this G_α and other GTPases prompts some thoughts about how the G_α switch turns transmembrane signals on and off.

The G_α proteins switch between inactive and active conformational states bound to GDP and GTP, respectively (for review see refs 2 and 3). Receptors activate G_α by binding to the α·GDP·$\beta\gamma$ heterotrimer and catalysing release of GDP from the guanine-nucleotide-binding site of G_α; then GTP binds, triggering dissociation of the $\beta\gamma$ complex and freeing α·GTP to stimulate the effector. α·GTP subsequently inactivates itself by hydrolysing the bound GTP and returning to the α·GDP state.

Although $G_{t\alpha}$ is the first G_α to be seen in three dimensions, there are crystal structures for two members of the much larger GTPase superfamily: $p21^{ras}$ (refs 4 and 5), the product of the *ras* oncogene, and elongation factor EF-Tu (refs 6–9), a bacterial protein involved in protein biosynthesis. Despite a meagre 16–18 per cent identity in primary structure among the three proteins, their GTPase core domains bear a striking family resemblance in three dimensions — the β-sheet and α-helices are arranged identically, and the contacts with the bound GTP analogue, GTPγS, are similar. This result satisfies investigators who have used the $p21^{ras}$/EF-Tu fold as a framework for modelling relations of G_α structure to function[10–12]. Experiments with peptides, bacterial toxins, antibodies, crosslinking agents, and chimaeric and other mutant recombinant proteins have assigned G_α amino-acid residues to probable surfaces on G_α for contacting receptors, effectors and $\beta\gamma$ (for review see ref. 12). Residues so assigned form contiguous surfaces on the $G_{t\alpha}$, and these surfaces are in the predicted locations.



How an activated receptor (R*) may turn on a G protein, by inducing the helical domain (green) of a G_α to pivot away from the G_α core GTPase domain (blue). By binding to surface features (red), the receptor loosens a cooperative web of hydrogen bonds (yellow circles), causing the helical domain to pivot away from the guanine-nucleotide binding site and allowing free exchange of GDP for GTP. View based on $G_{t\alpha}$ structure (Fig. 1*b*, page 655).

Now comes the lovely surprise: a stretch of 113 amino acids in the new $G_{t\alpha}$ structure (conserved in G_α chains but missing in other GTPases) comprises a second domain, composed of α-helices and inserted into the core domain at a site corresponding to loop 2 of $p21^{ras}$. The $G_{t\alpha}$ three-dimensional structure strongly suggests that this second domain is centrally involved both in turning the G_α switch on (GDP release, followed by GTP binding) and also in turning it off (GTP hydrolysis).

With respect to how G_α gets turned on, a quick look at the $G_{t\alpha}$ structure tells the story (see the accompanying figure, and Fig. 1*b* on page 655). Between them, the helical domain and the core GTPase domain thoroughly envelop bound guanine nucleotide. Like a teapot lid, the helical domain restricts movement out of (and into) the guanine-nucleotide-binding pocket of the core domain. To promote replacement of GDP by GTP, the activating receptor must somehow pry the lid away from the pot. Noel *et al.* suggest that the receptor does so by loosening a cooperative web of hydrogen bonds and van der Waals contacts connecting residues in both the core and helical domains with one another and with the guanine ring of the bound nucleotide. Several amino-acid residues in this web around the guanine ring belong to G_α sequences previously identified as probable receptor-interacting regions. These regions (red) and residues that contribute to the web (yellow) are depicted in the figure here (for details see refs 1 and 12).

Note that $p21^{ras}$ and EF-Tu bind guanine nucleotides quite tightly without the benefit of a large extra helical domain inserted into loop 2. For some Ras-like GTPases, however, a class of separate proteins — guanine-nucleotide-dissociation inhibitors (GDIs) — serves to inhibit egress of bound guanine nucleotide[13]. Thus the G_α helical domain may act as an intrinsic GDI that is subject to regulation by hormone receptors.

The G_α turn-off reaction, GTP hydrolysis, also depends on the helical domain. In the $G_{t\alpha}$ structure, the guanidinium group of an arginine (R174, located in a linker sequence between the helical and core domains) is positioned close to the γ-phosphate of bound GTPγS, where it could stabilize a transition state of the GTPase reaction. Functional studies had already shown that the same arginine

1. Noel, J. P., Hamm, H. E. & Sigler, P. B. *Nature* **366**, 654–663 (1993).
2. Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **348**, 125–132 (1990).
3. Kaziro, Y., Itoh, H., Kozasa, T., Nakafuku, M. & Satoh, T. *A. Rev. Biochem.* **60**, 349–400 (1991).
4. Tong, L., Milburn, M. V., deVos, A. M. & Kim, S.-H. *Science* **245**, 243–244 (1989).
5. Pai, E. F. *et al. EMBO J.* **9**, 2351–2359 (1990).
6. Jurnak, F. *Science* **230**, 32–36 (1985).
7. La Cour, T. F. M., Nyborg, J., Thirup, S. & Clark, B. F. C. *EMBO J.* **4**, 2385–2388 (1985).
8. Berchtold, H. *et al. Nature* **365**, 126–132 (1993).
9. Kjeldgaard, M., Nissen, P., Thirup, S. & Nyborg, J. *Structure* **1**, 35–49 (1993).
10. Masters, S. B., Stroud, R. M. & Bourne, H. R. *Protein Engng.* **1**, 47–54 (1986).
11. Holbrook, S. R. & Kim, S.-H. *Proc. natn. Acad. Sci. U.S.A.* **86**, 1751–1755 (1989).
12. Conklin, B. R. & Bourne, H. R. *Cell* **73**, 631–641 (1993).
13. Ullrich, O. *et al. J. biol. Chem.* **268**, 18143–18150 (1993).
14. Landis, C. A. *et al. Nature* **340**, 692–696 (1989).
15. Lyons, J. *et al. Science* **249**, 655–659 (1990).
16. Weinstein, L. S. *et al. New Engl. J. Med.* **325**, 1688–1695 (1991).
17. Abood, M. E., Hurley, J. B., Pappone, M.-C., Bourne, H. R. & Stryer, L. *J. biol. Chem.* **257**, 10540–10543 (1982).
18. Van Dop, C., Tsubokawa, M., Bourne, H. R. & Ramachandran, J. *J. biol. Chem.* **259**, 696–698 (1984).
19. Markby, D. W., Onrust, R. & Bourne, H. R. *Science* **262**, 1895–1901 (1993).

plays an essential role in the turn-off reaction in other G_α chains: human endocrine tumours, in the pituitary and elsewhere, are caused by GTPase-inhibiting mutations in which other amino acids are substituted for the cognate arginine in the α-subunit (α_s) of G_s (refs 14–16), the G protein responsible for hormonal stimulation of adenylyl cyclase; cholera toxin causes lethal diarrhoea by inhibiting GTP hydrolysis by α_s in gut mucosa, and shuts off the GTPase of $G_{t\alpha}$ by ADP-ribosylating R174 (refs 17 and 18).

The actual catalytic side chain in G_α — a general base responsible for persuading an intervening water molecule to mount a nucleophilic attack on the γ-phosphate of GTP — remains elusive. The structure does not point to a GTPase-catalysing role for glutamine 200, a residue that is conserved in all G_α chains and is essential for GTP hydrolysis, as shown by point mutations of the cognate glutamine codon in several G_α genes[3]. As an alternative candidate for the key general base, Noel *et al.* nominate a nearby residue, glutamate 203, whose side chain could pivot into the right position to activate an appropriately positioned water molecule.

Observations in my laboratory show biochemically that the helical domain and the key arginine are critical in turning off G_α (ref. 19). We used two purified recombinant proteins encoded by different parts of the α_s gene. A protein called Gail (for G_α-insert-1) is composed of a sequence of 141 α_s residues, corresponding to the helical domain of $G_{t\alpha}$ and the linker region that contains R174 (R201 in α_s); a protein called Ralph ('Ras-like') includes the core GTPase domain of α_s, encoded by sequences spliced together at the site from which Gail has been excised. Ralph binds GTPγS and stimulates adenylyl cyclase, but does not hydrolyse GTP; addition of Gail protein confers on Ralph the ability to hydrolyse GTP, at a rate even faster than that of the holoprotein, α_s. Gail in which cysteine replaces R201 (as in the mutant α_s of pituitary tumours) does not stimulate GTP hydrolysis.

Taken together, the two new papers[1,19] make a compelling argument that the helical domain, Gail, serves as a built-in, intrinsic GTPase-activating protein (GAP) for G_α, analogous to the extrinsic GAPs that are required for rapid GTP hydrolysis in p21^{ras}, EF-Tu and other members of the GTPase superfamily. Similarly, Gail acts as an intrinsic, built-in GDI for G_α, as described above. During evolution, a borrowed Gail DNA sequence was probably grafted into an ancient GTPase gene at a site corresponding to loop 2 of a primordial Ralph domain. Mutation and selective pressure seem to have sculpted the new domain into an intrinsic GDI–GAP combination, which now helps to turn on and off the versatile G_α switches that transduce extracellular signals into appropriate cellular responses. □

Henry R. Bourne is in the Departments of Pharmacology and Medicine, and the Cardiovascular Research Institute, University of California School of Medicine, San Francisco, California 94143, USA.

GEOPHYSICS

Core models ring true

T. Guy Masters

THE inner core, the solid innermost part of our planet, is a mysterious place. Although it is only slightly smaller than the Moon, it comprises less than 1 per cent of the Earth's volume. Consequently, the sampling of the inner core by seismic waves is poor, but the available data show very odd wave-propagation characteristics. The anomalous behaviour also extends to those low-frequency modes of elastic–gravitational oscillation ('ringing') of the whole Earth that have some sensitivity to the inner core. On page 678 of this issue[1], Tromp attempts to explain the observed oscillations with a simple model of elastic anisotropy.

In Tromp's model, seismic velocities are a function of the direction of wave travel. Roughly speaking, the model requires that the inner core behave like a single crystal with cylindrical symmetry about an axis aligned with the rotation axis of the Earth. It will be a challenge to provide a quantitative explanation of the phenomenon.

Tromp's paper is not the first to suggest this kind of model; the subject has had a chequered history. The anomalous behaviour of modes of free oscillation which have most of their energy in the core was first recognized in 1981 (ref. 2), but it was several years before enough modes had been analysed to give useful constraints on Earth structure[3,4]. The data are compatible with a structure within the core that is dominantly axisymmetric and has an effect equivalent to that of an anomalously large ellipticity of figure. Some early models explained the observations by putting lateral variations of seismic velocity[3] in the outer core, even though it was recognized that such structures would be difficult to maintain in an inviscid fluid. Lateral variations in velocity in the inner core or variations in topography in the inner-core boundary were also considered but had to be extremely large in magnitude to explain the data even partially[3,4]. The reason is simply that some of the anomalous oscillations have very little energy in the inner core (some less than 2 per cent). This is what makes it so remarkable that anisotropy in the inner core can explain the observations.

The first anisotropic models of inner-core structure were proposed by researchers at Harvard in 1986, although new observational constraints were added: the travel times of seismic body waves crossing the inner core[5,6]. (It had been recognized in 1983 that compressional waves that travel parallel to the spin axis are significantly faster, arriving about 2 seconds earlier, than waves that travel in the equatorial plane[7].) These models did not provide a quantitative fit to both the mode data and travel-time data, but they were presented to demonstrate the viability of inner-core anisotropy as the most plausible explanation of the anomalous signals. The model designed to fit the mode data significantly overpredicted the observed travel-time anomaly. Moreover, it had most of the anisotropy concentrated right at the surface of the inner core, again at variance with the travel-time observations[8].

Much more complicated anisotropic models were later designed to provide an adequate fit to all the data then available[9], but new observational studies of both body waves and free oscillations continued to provide additional constraints. By 1992, more than 20 anomalous modes had been analysed, and it was found that existing models of inner-core anisotropy provided only a poor fit to the full data set[10]. The body-wave studies continued to provide support for anisotropy[11], but of a much smaller magnitude than originally proposed and far too small to explain the mode data. Some authors (including myself) were led to speculate that inner-core anisotropy could not be the source of anomalous mode behaviour, although plausible alternatives were lacking.

The situation changed dramatically in 1992. At distances of 145–155° from the epicentre of an earthquake, it is possible to see two body-wave arrivals on the same seismogram, one of which passes just above the inner core (this is called BC) and the other of which turns 100–300 km inside the inner core (DF). The differential travel times between the two waves make ideal data for studying the top of the inner core, as the waves have almost identical paths through the outer core and mantle. Before 1992, no studies of differential times had included measurements from rays travelling nearly parallel to the spin axis. The differential times for such rays were found to be extremely anomalous[12], an important observation that has now been corroborated by other workers (X. Song and D. V. Helmburger,

manuscript in preparation). Clearly, this is much more compatible with the kind of anisotropy needed to explain the modes. Indeed, Tromp demonstrates that it is possible to find models that provide convincing fits to both mode and travel-time data sets.

Not surprisingly, there are still some puzzles. For example, all records that show anomalous DF–BC differential times have DF waveforms that are unusually small in amplitude and unusually complicated (this might explain why they have not been noticed before). As yet there is no satisfactory explanation for this. Also, a few modes are still poorly fitted and require further study. Overall, though, the hypothesis that strong inner-core anisotropy exists near the top of the inner core now seems to be on firm ground. Perhaps the biggest remaining puzzle is its cause. A major cause of anisotropy in the Earth's mantle is thought to be alignment of anisotropic crystals by convective flow. It is extremely likely that the inner core is undergoing solid-state convection and, until recently, it was accepted that the crystal structure of iron at inner-core pressures would be anisotropic, so making this a good candidate explanation[13]. In the past year, revisions to the high-pressure phase diagram of iron have been suggested, and it is not yet clear whether inner-core material can be anisotropic. Another possibility is that anisotropic fabric has been frozen into the inner core as it solidifies from the outer core.

Strong anisotropy is just one of many unusual properties of the inner core to be explained (others include strong and possibly frequency-dependent attenuation, and an unusual Poisson ratio) and provides one more clue to the origin and evolution of this most enigmatic part of the Earth. □

T. Guy Masters is at the Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093, USA.

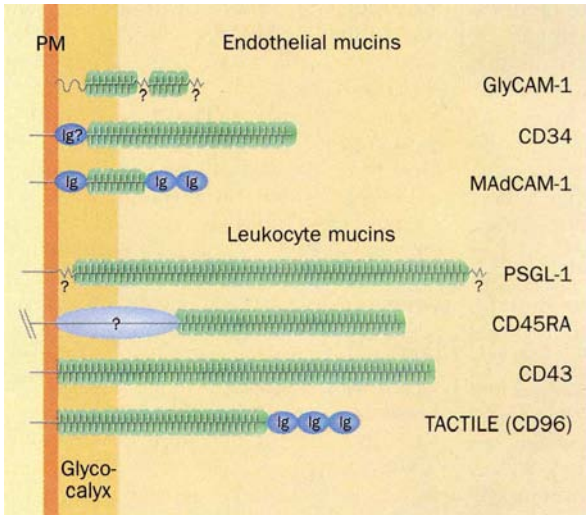
1. Tromp, J. *Nature* **366**, 678–681 (1993).
2. Masters, G. & Gilbert, F. *Geophys. Res. Lett.* **8**, 569–571 (1981).
3. Ritzwoller, M., Masters, G. & Gilbert, F. *J. geophys. Res.* **91**, 10203–10228 (1986); **93**, 6369–6396 (1988).
4. Giardini, D., Li, X.-D. & Woodhouse, J. H. *Nature* **325**, 405–411 (1987); *J. geophys. Res.* **93**, 13716–13742 (1988).
5. Morelli, A., Dziewonski, A. M. & Woodhouse, J. H. *Geophys. Res. Lett.* **13**, 1545–1548 (1986).
6. Woodhouse, J. H., Giardini, D. & Li, X.-D. *Geophys. Res. Lett.* **13**, 1549–1552 (1986).
7. Poupinet, G., Pillet, R. & Souriau, A. *Nature* **305**, 204–206 (1983).
8. Shearer, P. M., Toy, K. M. & Orcutt, J. A. *Nature* **333**, 228–232 (1988).
9. Li, X.-D., Giardini, D. & Woodhouse, J. H. *J. geophys. Res.* **96**, 551–577 (1991).
10. Widmer, R., Masters, G. & Gilbert, F. *Geophys. J. int.* **111**, 559–576 (1992).
11. Shearer, P. M. & Toy, K. M. *J. geophys. Res.* **96**, 2233–2247 (1991).
12. Creager, K. C. *Nature* **365**, 309–314 (1992).
13. Jeanloz, R. & Wenk, H.-R. *Geophys. Res. Lett.* **15**, 72–75 (1988).

CELL ADHESION

Mucins in the mainstream

Yoji Shimizu and Stephen Shaw

THE selectin family of carbohydrate-binding proteins has a singular role in initiating the 'adhesion cascade' by which leukocytes move from the blood stream into tissue[1]. Two molecular partners for selectins have already been identified[2,3], and this week we have two more[4,5]. All are the heavily *O*-glycosylated proteins known as mucins, and mucin versatility is illustrated by Berg and colleagues on page 695 of this issue[4] — their findings suggest that a mucin domain and an integrin-binding domain, present in a single molecule called MAdCAM-1, cooperate in adhesion.



Some of the mucins present on endothelial cells and leukocytes. The immunoglobulin (Ig) domains are indicated; the potential Ig domain in CD34 is smaller than a prototypic Ig domain. A question mark indicates a region without homologies to known structures. The overall length of molecules is estimated from known properties of homologous molecules, and the surface of a typical glycocalyx is shown at 10 nm (100 Å) above the plasma membrane (PM).

The immunological incentive for studying selectins is their involvement in the movement of circulating cells such as lymphocytes and neutrophils to various places in the body[1]. For example, L-selectin helps in the migration of lymphocytes into peripheral lymph nodes and sites of chronic inflammation, as well as neutrophil influx into acute inflammatory sites. The critical interaction that allows leukocytes to leave the circulatory highway is with endothelial cells, whether they be the specialized high endothelial venules (HEV) involved in lymphocyte migration into lymphoid organs or activated vascular endothelial cells at a site of injury.

Although L-selectin was initially thought to be a homing receptor which itself could mediate a specific pattern of lymphocyte migration, it is now seen as one element in achieving specific migration. The consensus model now has it that selectins initiate weak tethering to the endothelium, which is followed by activation of strong shear-resistant adhesion mediated by integrins[6,7]. The specificity of the final process depends, combinatorially, on the specificity at each of its three steps, and therefore diversity in selectin ligands contributes in a multiplicative fashion to specificity of the entire cascade.

During the past 18 months three L-selectin ligands and one P-selectin/E-selectin ligand have been cloned, all of which are mucins. The first L-selectin ligand was GlyCAM-1, which contains two regions rich in serine/threonine that are extensively *O*-glycosylated[2]. The requirement for GlyCAM-1 sulphation for function[8], and the almost exclusive expression of GlyCAM-1 in lymph node HEV, fulfilled expectations that here was an L-selectin ligand. But L-selectin also recognizes a molecule of M_r 90K, which has been identified as the sialomucin CD34 (whose only previous claim to fame was as a useful marker for pluripotent stem cells in the bone marrow). CD34 is also expressed on endothelium but, unlike GlyCAM-1, its expression is not restricted to lymph nodes.

MAdCAM-1 was not expected to be an L-selectin ligand, because it is found primarily on endothelium in mucosal tissue and participates in specific homing of lymphocyte subsets to mucosal tissue[9]. Functional studies demonstrated that MAdCAM-1 is a ligand for the $\alpha4\beta7$ integrin, which had previously been implicated in lymphocyte homing to the gut[10]. Not surprisingly, cloning of the MAdCAM-1 gene revealed the presence of two amino-terminal immunoglobulin-like domains similar to domains in two other integrin counter-receptors, ICAM-1 and VCAM-1 (ref. 9). What was surprising was that MAdCAM-1 also contains a mucin-like domain.

Berg and co-workers[4] have pursued this serendipitous finding and now show that MAdCAM-1 is also a ligand for L-selectin. Like other selectin ligands, MAdCAM-1 supports the low-affinity interaction of L-selectin$^+$ $\alpha4\beta7^-$ cells under shear (fluid flow) that manifests itself as

'rolling' of cells along a MAdCAM-1-coated surface. So MAdCAM-1 has the structural requirements to contribute to both the rolling and strong adhesion steps of the adhesion cascade for the same cell. Although it has not been formally demonstrated, this concept has tremendous appeal because it is reminiscent of the multiple domains found in other molecules (for example the adhesive domains in extracellular matrix proteins such as fibronectin, and the SH2/SH3/kinase domains in tyrosine kinases).

Given the multiplicity of L-selectin ligands, it might seem that the protein cores of these molecules would be functionally interchangeable or degenerate. But a report of the cloning of PSGL-1, a ligand for P-selectin (and E-selectin)[5], shows that this is not the case. P-selectin-dependent binding requires expression of the PSGL-1 protein core and cannot be replaced by expression of a similar mucin core protein, CD43. So mucins are ligands not only for L-selectins but also for other selectins, and the protein core helps dictate the specificity of the interaction.

What are mucins and why are they selectin ligands? The defining characteristics of GlyCAM-1, MAdCAM-1, CD34, PSGL-1 and other mucins (see figure) is a high local density of *O*-linked sugars[11]. Prototypic mucins like CD43 (sialophorin, leukosialin) consist primarily of such regions. But in most cases mucin domains of variable sizes are interspersed with other structural motifs, giving rise to potential combinatorial effects. Notably, in three of the cases illustrated in the figure, mucin domains occur with immunoglobulin domains, meaning that the adhesive collaboration proposed for MAdCAM-1 may be a general feature of these molecules.

The dense array of *O*-linked side chains in mucins has at least two important structural implications. The first is extended structure; the average extension per amino acid is predicted to be 2.5 Å, making many mucins long enough for them to gain exposure above the glycocalyx (the glycoprotein and polysaccharide covering that surrounds cells). The second is optimal exposure and high multiplicity of the terminal sugars. These two features make mucins powerful two-edged swords: they are both anti-adhesive and pro-adhesive. By virtue of their negative charge and extended configuration, mucins act as a repulsive barrier around a cell. However, when an opposing cell has specific receptors for the mucins' sugars, adhesion supplants repulsion; such rapid interactions of lectins with mucins serve as the molecular brake which slows down leukocytes under conditions of shear[12], and initiate the first embrace between sperm and unfertilized egg[13].

Mucins play numerous tricks to increase their versatility in regulating adhesion. First, expression of the core protein is regulated, as illustrated by GlyCAM-1. Second, mucin domains are subject to alternative splicing, for example in CD45, allowing mucin function to be regulated independently of the rest of the molecule. Third, the structure of the *O*-linked side chains is regulated by cell-specific differences in glycosyltransferase expression. For instance, at some sites MAdCAM-1 fails to function as a L-selectin ligand because it is not decorated with the critical carbohydrates necessary for recognition. And fourth, there are the tricks related to anchorage to the plasma membrane: GlyCAM-1 lacks a classic transmembrane region, and its attachment to the endothelial cell surface occurs through an incompletely defined mechanism which facilitates shedding of GlyCAM-1; others undergo cleavage from their transmembrane tails.

Like most findings in cell adhesion these days, the selectin/mucin story is a glorious illustration of regulatory complexity begetting exquisite specificity. The tasks in hand are understanding fully how that specificity comes about, and defining the functions of the many other mucins. □

Yoji Shimizu is in the Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. Stephen Shaw is in the Experimental Immunology Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.

1. Lasky, L. A. *Science* **258**, 964–969 (1992).
2. Lasky, L. A. *et al. Cell* **69**, 927–938 (1992).
3. Baumhueter, S. *et al. Science* **262**, 436–438 (1993).
4. Berg, E. L., McEvoy, L. M., Berlin, C., Bargatze, R. F. & Butcher, E. C. *Nature* **366**, 695–698 (1993).
5. Sako, D. *et al. Cell* **75**, 1179–1186 (1993).
6. Butcher, E. C. *Cell* **67**, 1033–1036 (1991).
7. Shimizu, Y., Newman, W., Tanaka, Y. & Shaw, S. *Immun. Today* **13**, 106–112 (1992).
8. Imai, Y., Lasky, L. A. & Rosen, S. D. *Nature* **361**, 555–557 (1993).
9. Briskin, M. J., McEvoy, L. M. & Butcher, E. C. *Nature* **363**, 461–464 (1993).
10. Berlin, C. *et al. Cell* **74**, 185–195 (1993).
11. Carraway, K. L. & Hull, S. R. *Glycobiology* **1**, 131–138 (1991).
12. Williams, A. F. *Nature* **352**, 473–474 (1991).
13. Gahmberg, C. G., Kotovuori, P. & Tontti, E. *APMIS Suppl.* **27**, 39–52 (1992).

PLANT EVOLUTION

Time for the angiosperms

Peter R. Crane

ANGIOSPERMS, the flowering plants, are the most diverse group of plants on the planet and in the past 25 years there has been extraordinary progress in understanding their early evolution. No single idea has been more influential in guiding research than the conventional view that angiosperms first appear in the fossil record about 130 million years ago and underwent their initial radiation through the mid- and late Cretaceous. This basic tenet of palaeobotanical research is now challenged by Bruce Cornet, who in a paper in *Modern Geology*[1] describes the discovery of putative angiosperm fossils in late Triassic rocks from eastern North America. If these specimens are angiosperms they would be more than 200 million years old.

From even the most casual overview of the plant fossil record, it is obvious that a considerable diversification of angiosperms occurred through the early and mid-Cretaceous, but for many years this interpretation was confused by records of supposed pre-Cretaceous angiosperms and the apparent diversity and modernity of Cretaceous taxa. Together, these factors fuelled speculation about a long period of 'cryptic' angiosperm evolution before their appearance in the fossil record. In 1960 an influential paper by Scott, Barghoorn and Leopold[2] questioned, and in some cases debunked, many of the pre-Cretaceous angiosperm claims. Comparison of fossil angiosperms with their supposed living relatives also called into question the assumed diversity and modernity of Cretaceous taxa[3], and with new information on fossil angiosperm pollen and leaves[4–6] made a long period of cryptic evolution seem unlikely.

This view has held sway ever since. But the possibility of a substantial pre-Cretaceous record has been raised again, from molecular-clock calculations[7,8] and also cladistic hypotheses that link angiosperms with Bennettitales (an extinct group of Triassic–Cretaceous seed plants) and Gnetales[9,10]. Because both of these potential angiosperm sister groups were present by the Triassic, the implication is that the lineage leading to angiosperms (stem angiophytes[11]) must have already diverged by at least 230 million years ago.

Enter Cornet, who in a series of papers has vigorously resurrected the possibility of the existence of pre-Cretaceous angiosperm-like plants. His evidence is dispersed fossil pollen[12], including that of the 'Crinopolles' group[13], reproductive structures attached to *Sanmiguelia*-like leaves[14] and now a dicot-like leaf and two putative angiosperm reproductive structures from the late Triassic (late Carnian[1]).

These new specimens are preserved in black lacustrine shales from the Cow Branch Formation (Newark Supergroup) of the Dan River/Danville Basin of North Carolina and Virginia. The leaf is small and clearly shows reticulate venation

which, although it is indeed strikingly angiosperm-like, also resembles that of certain dipteridaceous ferns that are common in the Newark Supergroup. The reproductive organs are more difficult to interpret and the critical details necessary to establish the presence of angiosperm carpels, rather than the cupules or ovules of other seed plants, are not preserved. In my view, the case for the angiosperm affinities of both the leaf and reproductive structures is inconclusive.

So is the current view of angiosperm macroevolution totally secure? Probably not, but the major mid-Cretaceous diversification is indisputable and there are limits to the kind of modifications that might be expected. In particular, because distinctive triperturate pollen is diagnostic of a major clade of dicotyledons (eudicots) consisting of more than 70 per cent of extant angiosperm species, and because such grains have an excellent fossil record, it is unlikely that members of this group existed before about 120 million years ago. But although we do not yet have unequivocal evidence for pre-Cretaceous angiosperms, we do have accumulating evidence of diverse Triassic, Jurassic and Cretaceous plants (such as *Sanmiguelia*) that cannot be placed conveniently in any of the currently recognized 'gymnosperm' groups[15].

Palaeobotanists need to be open to the possibility that fossils such as these could be stem-angiophytes. Particularly intriguing are Cornet's descriptions of the exceedingly angiosperm-like Triassic 'Crinopolles' pollen grains for which the parent plants are so far unknown[16]. Together, the new discoveries are a useful reminder that we know less than we think about the plant diversity of 200 to 100 million years ago, and highlight the fact that we are still missing groups that may be key elements in understanding seed-plant phylogeny and the origin of flowering plants. □

Peter R. Crane is in the Center for Evolutionary and Environmental Biology, Field Museum, Chicago, Illinois 60605, USA.

1. Cornet, B. *Mod. Geol.* **19**, 81–99 (1993).
2. Scott, R. A., Barghoorn, E. S. & Leopold, E. B. *Am. J. Sci.* **258-A**, 284–299 (1960).
3. Dilcher, D. L. *Bot. Rev.* **40**, 1–157 (1974).
4. Doyle, J. A. *J. Arnold Arbor.* **50**, 1–35 (1969).
5. Muller, J. *Biol. Rev.* **45**, 417–450 (1970).
6. Doyle, J. A. & Hickey, L. J. in *Origin and Early Evolution of Angiosperms* (ed. Beck, C. B.) 139–206 (Columbia University Press, New York, 1976).
7. Martin, W. A. *et al. Nature* **339**, 46–48 (1989).
8. Crane, P. R., Donoghue, M. J., Doyle, J. A. & Friis, E. M. *Nature* **342**, 131–132 (1989).
9. Crane, P. R. *Ann. Mo. bot. Gdn* **72**, 716–793 (1985).
10. Doyle, J. A. & Donoghue, M. J. *Bot. Rev.* **52**, 321 (1986).
11. Doyle, J. A. & Donoghue, M. J. *Paleobiology* **19**, 141–167 (1993).
12. Cornet, B. & Habib, D. *Rev. Palaeobot. Palynol.* **71**, 269–294 (1992).
13. Cornet, B. *Palaeontographica Abt.* B**213**, 37–87 (1989).
14. Cornet, B. *Evolut. Theory* **7**, 231–309 (1986).
15. Crane, P. R. *Taxon* **36**, 778–779 (1987).
16. Doyle, J. A. & Hotton, C. L. in *Pollen and Spores: Patterns of Diversification* (eds Blackmore, S. & Barnes, S. H.) 169–195 (Clarendon, Oxford, 1991).

QUANTUM OPTICS

Atomic trampolining

C. J. Foot

CAESIUM atoms have been made to bounce up and down nearly 10 times above a curved glass surface in a novel experiment at the Ecole Normale Supérieure in Paris (C. G. Aminoff *et al.*, *Phys. Rev. Lett.* **71**, 3083–3086; 1993). The surface is a mirror where matter is reflected by light, in this case the strong repulsive potential produced by an evanescent wave extending from the glass into the vacuum.

In previous work with flat surfaces, only a few bounces on the atomic trampoline were observable, but with a curved surface the atoms do not spread out because they are focused. The number of bounces is limited only by scattering processes. As well as storing the atoms for long periods, the authors hope to see new phenomena associated with the wave-like nature of the atoms, analogous to the interference effects that occur for light. There is much current interest in making mirrors, beam-splitters and other optical elements for matter waves so that interferometers can be constructed and precise measurements made.

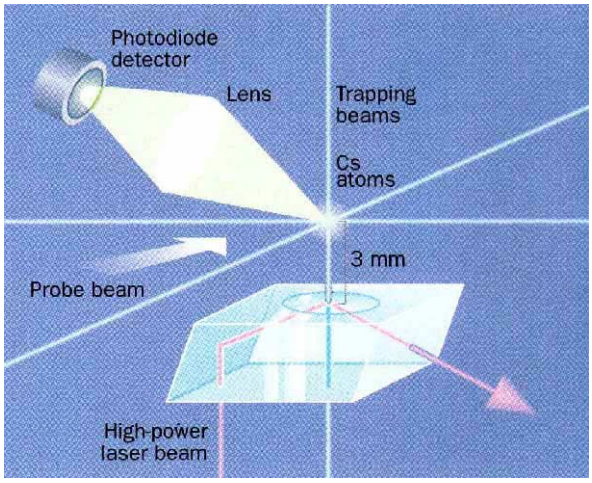


FIG. 1 The atomic trampoline devised by Aminoff and colleagues. Caesium atoms are collected at the intersection of the counter-propagating trapping beams and slowed to very low velocities. When these beams are shut off, the atoms drop under gravity and bounce on the evanescent wave at the curved surface of the glass block. The number of atoms returning is measured by turning on a probe beam and observing the fluorescence on a photodiode. From many runs with different probing times, several bounces of the atoms can be seen (Fig. 2).

An evanescent wave is formed when a beam is totally internally reflected within a medium, as shown in Fig. 1. In their experiments, Aminoff and colleagues directed a powerful laser beam into a glass prism so that, on its second internal reflection, it was incident at an acute angle on the upper face at a point where the surface had been polished concave (radius of curvature 20 mm). Although all of the beam is reflected, the electric field associated with the light cannot be zero in the vacuum just on the other side of the boundary — such a discontinuous change would give infinite field gradients. For a valid solution of Maxwell's equations, there must be something on the vacuum side of the interface. This is the evanescent wave. Its associated fields fall away exponentially over a distance roughly equal to the wavelength of the light, giving a very high intensity gradient.

When the frequency of the laser is slightly above the atomic resonance frequency, the radiation force tends to push atoms away from high-intensity regions, and the evanescent wave acts like a thin reflective coating on the surface. However, if atoms are excited by the light and emit spontaneous photons, the recoil gives the atom a small random kick so that the 'reflection' is no longer purely specular. This process destroys the coherence of the matter waves and would prevent any future observation of interference effects. Fortunately, as the laser frequency is moved away from the atomic resonance, the rate of spontaneous emission falls off faster than the repulsive force in the evanescent wave. For the intensity used — a 0.8-watt beam focused down to a spot 1 mm in diameter on the surface — the laser frequency could be 10 GHz above the resonance while still giving a sufficiently strong force. This made the rate of spontaneous emission only 1 photon per 11 bounces.

The atoms to be dropped on this small mirror were first collected by a magneto-optical trap in which the combined effects of the forces from three orthogonal pairs of counter-propagating laser beams (shown in Fig. 1) and a magnetic field (coils not shown in the figure) both cool the atoms and compress them into a small dense cloud. This is a convenient and widely used source of ultracold atoms which has been described in previous News and Views articles (C. J. F. and A. Steane, *Nature* **344**, 490; 1990 and **351**, 521; 1991). After the trap had been loaded, the magnetic field was switched off and the intensity of all the trapping beams reduced to obtain temperatures of

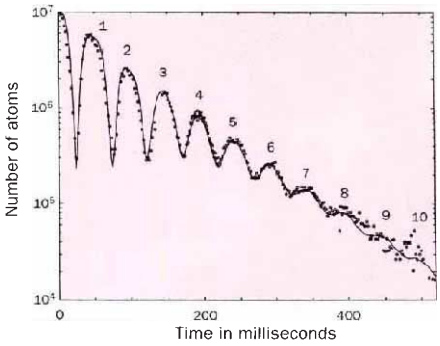


FIG. 2 Numbers of atoms bouncing back (from the paper by Aminoff *et al.*).

only 5 μK. When the trapping beams were shut off completely, the cloud held typically 10^7 ultracold atoms which then dropped a height of 3 millimetres to the mirror.

After bouncing, the atoms were observed by switching on a probe laser beam and detecting the fluorescence with a photodiode. This means of detection disrupts the bouncing atoms, and so the experimenters made many runs with different probing times to follow the atoms over many bounces (see Fig. 2). In each run the losses were initially high because the mirror was smaller than the atomic cloud, but after the first few bounces, 73 per cent of the atoms returned after each bounce. In future, it should be possible to reduce the limiting processes (such as scattering by stray light and background gas, or the spontaneous emission described above) and so allow many more bounces. This would give a very long time during which the atoms are freely falling, being perturbed by the light only for short, well defined periods, and so would allow very precise measurements to be made, as in a hydrogen maser.

Of course, as the authors themselves say, the purpose of these experiments is "not merely to demonstrate juggling with unusually small balls". One can adopt a wave-like description of the atom cavity in which de Broglie waves are reflected coherently by the mirror; this is rather like light travelling back and forth between the two mirrors of a Fabry–Perot etalon but with gravity playing the role of one of the mirrors. Gravity also causes the potential energy of the atoms to vary throughout the cavity, which therefore corresponds to a medium of varying refractive index in optics. With these modifications, the spatial distribution of the atomic wavefunction can be calculated, much as one would for gaussian light beams. The observation of this and other matter-wave phenomena analogous to those seen in lasers would be truly fascinating. □

C. J. Foot is in the Atomic and Laser Physics Department, Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, UK.

ASTROPHYSICS

Chaos in interstellar clouds

Stephen Lepp

THE discovery by Le Bourlot, Pineau des Forêts, Roueff and Schilke[1] that interstellar chemical models have bistable solutions is both one of the most exciting new developments in astrochemistry and potentially one of the most disappointing. The authors have modelled interstellar chemical abundances and found that for a range of densities and temperatures there are two stable equilibrium solutions. Which of these solutions is reached depends on the initial conditions. This bifurcation of a single steady-state solution is exciting because it suggests that the chemical abundances in an interstellar cloud may hold some indication of the history of the cloud, and because of the relation of these bifurcations to chaos. It is potentially disappointing because if these chemical abundances are chaotic, it may be impossible to infer physical conditions from them.

Chemical chaotic systems have been known for some 30 years, the first being discovered by Boris Belousov who became intrigued by oscillations in a particular mix of chemicals. His paper was rejected and he was told[2] that his "supposed discovery" was impossible. Most chemists felt that it violated the second law of thermodynamics, and that all reactions would settle down to the lowest energy state. Years later another Russian, Anatol Zhabotinskii, learned of the reaction and managed to publish the results of a careful investigation[2]. As is now well known, it is because the Belousov–Zhabotinskii reaction is so far from thermodynamic equilibrium that it can exhibit such strange behaviour. Because the interstellar medium is also far from thermodynamic equilibrium it is perhaps not surprising that it shows bistable solutions.

What is surprising, and what makes this result more than an intellectual exercise, is that the two states coexist in the model developed by Le Bourlot *et al.* for a range of densities near 10^4 molecules per cubic centimetre and with temperatures ranging from 10 to 40 K, just the physical conditions expected for most interstellar clouds. The two states can be characterized by their very different degrees of ionization (that is, the ratio of ionized to neutral species). The high-ionization phase is roughly an order of magnitude more ionized than the low-ionization phase for the same density and temperature.

Interstellar clouds are very cold, between 10 and 20 K, and there would be little chemistry if it were not for the nonthermal input from cosmic rays. Cosmic rays ionize molecular hydrogen, producing both H^+ and H_2^+. The H_2^+ immediately reacts with H_2 to produce H_3^+. From here a large number of ion–molecule reactions account for the variety of species. At very low densities, interstellar gas is highly ionized because the ionization depends on the density of the gas while the recombination rate depends on the square of the density. At high fractional ionization, the molecular ions disappear through rapid dissociative recombinations and H^+ is the predominant hydrogen ion. At high densities, the fractional ionization is low and H_3^+ predominates. The bistable solution occurs at densities and temperatures where the high-ionization and low-ionization states can coexist.

The location of the bistable region is also sensitive to some of the other model parameters. For example, S^+ is one of the major ions in the cloud and so sulphur contributes to the concentration of free electrons, which in turn determines if and where a bistable region occurs.

The bifurcation also depends critically on the assumed value of the rate constant for dissociative recombination of H_3^+

$$H_3^+ + e^- \rightarrow H_2 + H$$
$$\rightarrow H + H + H$$

Le Bourlot and colleagues adopted a rate constant extrapolated from a measured[3] value of 1.5×10^{-7} cm^3 s^{-1} at 300 K, which may be too large. Changing the value of this constant changes the densities for which the two solutions overlap. For values smaller than 2.5×10^{-8} cm^3 s^{-1}, the bistable region disappears and one again has a single stable solution throughout the region.

This sensitivity is particularly tantalizing because of the range of measured rate constants, with merged beam experiments[4] producing values as low as 2×10^{-8} cm^3 s^{-1} and flowing afterglow measurements possibly giving still lower values[5]. Although the most recent measurements[6] seem to be converging to higher values, this study underscores the importance of this rate constant to interstellar chemistry and the need to determine its value.

Although the bistability is characterized by the degree of ionization, all the chemical abundances are affected. One particularly interesting example is the ratio of atomic carbon to carbon monoxide, which differs by more than an order of magnitude between the high-ionization phase, where C/CO is greater than 0.1, and the low-ionization phase, where C/CO is less than 10^{-2}.

This bistability may help to settle a long-standing problem in interstellar chemistry. Carbon on the surfaces of

interstellar clouds is ionized by starlight and is predominantly in the form of C^+. Deeper into the cloud, the light is attenuated by dust and one expects nearly all of the carbon to be in the form of CO. The problem is that observations[7] suggest that in certain regions the C/CO ratio stays high well into the clouds.

Some of these are photo-dominated regions, where the higher carbon abundance comes from a locally high radiation field, either through a group of nearby hot stars or through geometric factors which allow more photons to penetrate. The observations are more difficult in regions which are not photo-dominated because the gas is cooler and carbon does not emit as much radiation. Still, there have been suggestions of high neutral carbon abundances extending fairly deeply into some interstellar clouds. A follow-up study by the same group[8] suggests that the high-ionization phase may be responsible for these regions of high C/CO ratio.

The high C/CO ratio also means that there is more carbon available to form other species such as hydrocarbons. With proper modelling, other carbon-bearing molecules such as CH and CN can prove more reliable indicators of the atomic carbon abundance[8]. Ultimately, more and better observations of the carbon and carbon molecules are needed to understand the carbon chemistry in interstellar clouds. □

Stephen Lepp is in the Physics Department, University of Nevada, Las Vegas, Nevada 89154, USA.

1. Le Bourlot, J., Pineau des Forêts, G., Roueff, E. & Schilke, P. *Astrophys. J.* **416**, L88–90 (1993).
2. Scott, S. *New Scientist*, 53–59 (2 December 1989).
3. Canosa, A., Rowe, B., Mitchell, J. B. A., Gomet, J. C. & Rebrion, C. *Astr. Astrophys.* **248**, L19–21 (1991).
4. Hus, H., Yousif, F. B., Sen, A. & Mitchell, J. B. A. *Phys. Rev.* A**38**, 658–663 (1988).
5. Adams, N. G. & Smith, D. in *Astrochemistry* (eds Vardya, M. S. & Tarafdar, S. P.) 1–18 (Reidel, Dordrecht, 1987).
6. Larsson, M. *et al. Phys. Rev. Lett.* **70**, 430–433 (1993).
7. Genzel, R., Harris, A. I., Jattee, D. T. & Stutzki, J. *Astrophys. J.* **332**, 1049–1057 (1988).
8. Flower, D. R., Le Bourlot, J., Pineau des Forêts, G. & Roueff, E. *Astr. Astrophys* (in the press).

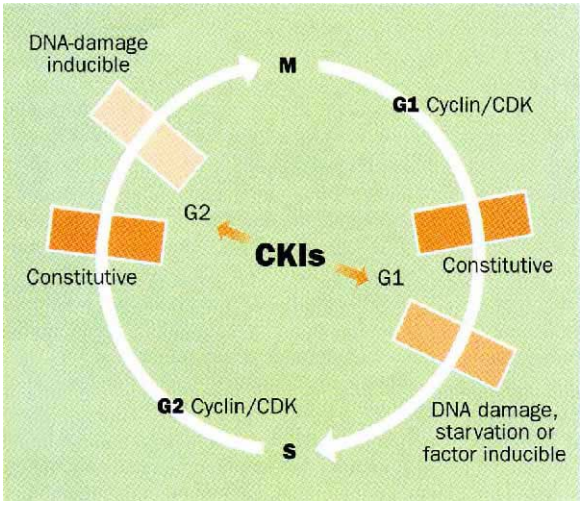
CELL CYCLE

Dams and sluices

Kim Nasmyth and Tim Hunt

CELLS in tissues have only three serious options in life — they can grow and divide, not grow but stay alive, or die by apoptosis. Tumours arise, we suppose, either by inappropriate growth and division or by cells failing to die when they should. Gaps in our knowledge of the behaviour of normal or malignant cells in human tissues have been partly filled by genetic analysis; this has revealed tumour-suppressing genes such as *p53* and *RB* that encode proteins involved in cell-cycle control. Onset of DNA replication and mitosis, the two key cell-cycle events, are triggered by successive waves of cyclin-dependent protein kinases (CDKs), comprising a cyclin regulatory subunit and a *CDC2*-family kinase subunit.

A flurry of papers[1–6], including three in this issue, now reveals the existence of hitherto unknown cell-cycle control elements, CDK inhibitors (CKIs). Such protein kinase inhibitors have long been known for the cyclic AMP-dependent protein kinase[7], but their involvement in cell-cycle regulation was not previously suspected. We believe they may be crucial components of the cell-cycle engine (see figure).



A simplified cell cycle with G1 and G2 cyclin kinase inhibitors (CKIs). The dark red blocks represent constitutive CKIs, the paler ones CKIs that can be induced under particular circumstances. These blocks to cell-cycle progression may be removed by proteolysis or phosphorylation.

In budding yeast, cell-cycle events are triggered by a single protein kinase subunit, Cdc28, which associates with a succession of different cyclins (Clns and Clbs). The programme is initiated in G1 phase by Cln3, which promotes accumulation of Cln1 and Cln2, leading to the cell-cycle commitment point known as Start. In S phase, Clb5 and Clb6 replace Cln1 and Cln2 and, at M phase, Cdc28's partners change to Clb1 and Clb2. The abundance of cyclins rises and falls as cells progress through the cycle. In mammals, more than one kinase subunit is implicated in cell-cycle control[8]. Thus progression from G1 to S phase involves Cdk4/cyclin D and Cdk2/cyclin E; S phase Cdk2/cyclin A; and M phase Cdc2/cyclin B. Cyclins D and E seem to be equivalent to Cln3 and Cln1/2 respectively, cyclin A to Clb5/6, and cyclin B to Clb1/2. Cln3 and cyclin D stand out, because their levels do not oscillate. They may act as 'initiator' cyclins that help to coordinate entry into the cell cycle with cell growth.

Cyclin oscillations are determined partly by transcriptional control of messenger RNA production and partly by specific proteolysis mediated by special ubiquitin ligases that turn on and off at particular times in the cell cycle. At its simplest, the programme of cyclin waves is in part an internal affair, normally tied to cell growth, and running more or less independently of the events triggered by cyclins. The best known example of this autonomy is the oscillation of maturation promoting factor (Cdc2/cyclin B) in frog eggs, whose sawtooth waves of kinase activity are independent of the nucleus[9].

But cyclin oscillations can be interrupted by external events. Remove nutrients from a yeast cell or growth factors from a mammalian cell and they arrest in a quiescent G1 state. How do such conditions stop the cell-cycle engine? The cell-cycle arrest seen in mating yeast is a good example, and sets a precedent for how CKIs may work. Small peptide pheromones, secreted by cells of the opposite mating type, bind to surface receptors which activate a MAP kinase cascade, which in turn phosphorylates a protein called Far1. Far1 can bind to and inactivate the Cdc28 protein kinase[10], causing G1 arrest because Far1 is normally only present in G1 (ref. 11). Far1 is thus the founder member of the CKI class. Other candidates include the yeast protein p40, which is a potent inhibitor of CDC28 kinases *in vitro*[12], and p16 from human cells which can inactivate Cdk4/cyclin D (ref. 4). CKIs are also implicated in the G1 arrest of mink lung epithelial cells by TGF-β; these cells contain Cdk2/cyclin E, which are complexed with a protein of M_r 27K (ref. 6).

Cell-cycle arrest also occurs in response to chromosome damage, and, in yeast, the *RAD9* gene is needed to delay mitosis when chromosomes are damaged by X-rays[13]. Inspired by the metaphor of a border crossing, Weinert and Hartwell coined the term 'checkpoint' to denote such surveillance mechanisms. This term has gained rapid popular acceptance, but like Humpty Dumpty, you can use it to mean almost anything you like in connection with cell-cycle control[14]. Checkpoints have been taken as those points in the cell cycle at which cells arrest in response to

damage[15], and as points of completion of processes that are being checked[16], as well as to denote the border guards themselves[14]. The term is probably here to stay, however, despite these ambiguities.

How does the Rad9 surveillance mechanism work? Cdc2 can be turned off by phosphorylation on tyrosine 15, which is important for controlling mitosis following a block to DNA replication in fission yeast[17]. But one can mutate tyrosine 15 to phenylalanine in budding yeast with no ill effects[18], so there must be another way of stopping the cell cycle besides phosphorylating Cdc2/Cdc28. Could it be CKIs?

Like *RAD9*, the tumour-suppressing gene *p53* is not required for normal cell growth or division, but it is required to shut down S phase after treatment with ultraviolet light[19–21]. Such inappropriate DNA replication in p53 mutant cells helps generate the mutations that lead to cancer. p53 is a nuclear protein that binds specific DNA sequences and is probably a transcription factor. If so, what genes does it control? The answer is: a member of the CKI family, a p21 protein that can bind to and inhibit a wide variety of cyclin-dependent kinases[1–3,5]. A rise in levels of p21 could well account for the ultraviolet-induced G1 arrest of normal human cells.

Cyclin kinase inhibitors may well have roles in the succession of cyclin waves during normal cell cycles. If so, how are these blocks removed or circumvented? One answer may be proteolysis. Several yeast *CDC* genes encode components of the proteolysis machinery. For example, *CDC34* and *UBC9* encode two different ubiquitin-conjugating enzymes, the first being needed for DNA replication, the second for mitosis[22]. Until now, it was suspected that these enzymes were needed for destroying cyclins — Cdc34 for destroying Cln1 and Cln2, and Ubc9 for destroying Clbs. We suspect that some of these *UBC* genes are needed to destroy CKIs, and not cyclins.

So we are led to a simple 'hydraulic' model for cell-cycle progression, according to which CDKs build up like water behind a CKI dam, first during G1. Once the dam is full, the surplus cyclins (possibly with help from others) trigger destruction or inactivation of the CKI — the dam bursts — and the cell is now irrevocably committed to S phase. But a second dam awaits at the G2/M transition, where we suppose a second kind of CKI lies in wait for the next CDK wave. When this CKI is overwhelmed, the cell enters mitosis: another irrevocable commitment. At the end of mitosis, the slate is wiped clean by the destruction of cyclins, the water evaporates, so to speak, whence it can return to the top of our imaginary mountain. How and when the dams are rebuilt, and whether the cycle is mainly interrupted by the heightening of old dams or the building of new ones, are questions for the future. □

Kim Nasmyth is at the Research Institute of Molecular Pathology, Dr Bohr-Gasse 7, A-1030 Vienna, Austria. Tim Hunt is at the ICRF Clare Hall Laboratories, South Mimms, Potters Bar, Hertfordshire EN6 3LD, UK.

1. El-Deiry, W. S. *et al. Cell* **75**, 817–825 (1993).
2. Harper, J. W., Adam, G. R., Wei, N., Keyomarsi, K. & Elledge, S. J. *Cell* **75**, 805–816 (1993).
3. Xiong, Y. *et al. Nature* **366**, 701–704 (1993).
4. Serrano, M., Hannon, G. J. & Beach, D. *Nature* **366**, 704–707 (1993).
5. Gu, Y., Turck, C. W. & Morgan, D. O. *Nature* **366**, 707–710 (1993).
6. Polyak, K. *et al. Genes Dev.* (in the press).
7. Cheng, H. C. *et al. J. biol. Chem.* **261**, 989–992 (1986).
8. Sherr, C. J. *Cell* **73**, 1059–1065 (1993).
9. Murray, A. W. & Kirschner, M. *Science* **246**, 614–621 (1989).
10. Peter, M., Gartner, A., Horecka, J., Ammerer, G. & Herskowitz, I. *Cell* **73**, 747–769 (1993).
11. McKinney, J. D., Chang, F., Heintz, N. & Cross, F. R. *Genes Dev.* **7**, 833–843 (1993).
12. Mendenhall, M. D. *Science* **259**, 216–219 (1993).
13. Weinert, T. A. & Hartwell, L. H. *Science* **241**, 317–322 (1988).
14. Hartwell, L. H. *Genetics* **129**, 975–980 (1991).
15. Murray, A. W. *Nature* **359**, 599–604 (1992).
16. Losick, R. & Shapiro, L. *Science* **262**, 1227–1228 (1993).
17. Enoch, T. & Nurse, P. *Cell* **60**, 665–673 (1990).
18. Nasmyth, K. *Curr. Opin. Cell Biol.* **5**, 166–179 (1993).
19. Kastan, M. B. *et al. Cell* **71**, 587–597 (1992).
20. Hartwell, L. *Cell* **71**, 543–546 (1992).
21. Painter, R. B. & Young, B. R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7315–7317 (1980).
22. Jentsch, D. A. *Rev. Genet.* **26**, 179–207 (1992).

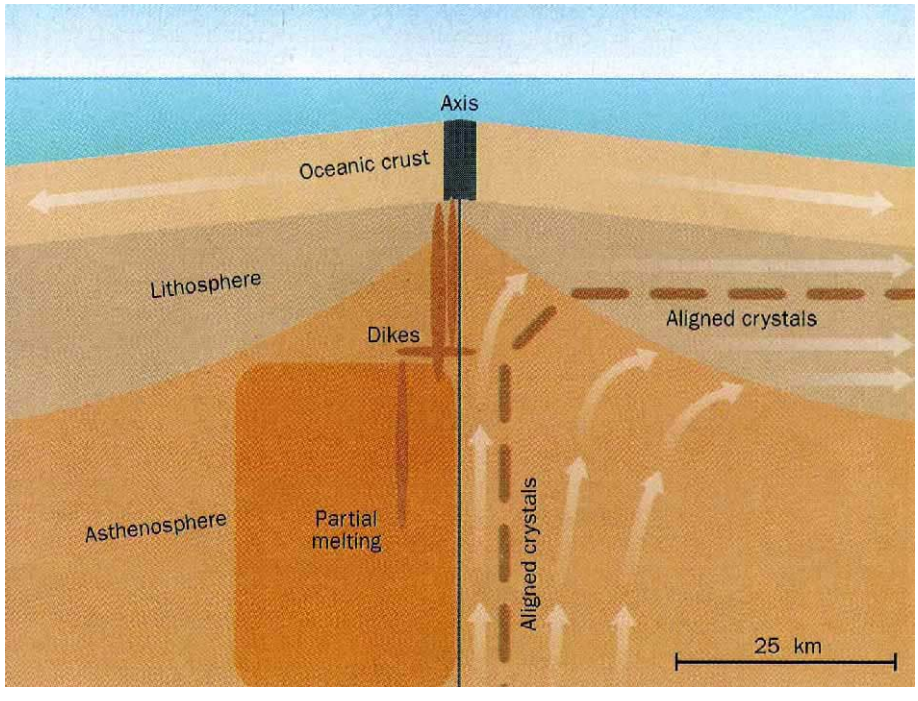
TECTONICS

Upwelling beneath ocean ridges

Norman H. Sleep

A KEY aspect of the theory of plate tectonics is the formation of oceanic crust at mid-ocean ridges. Originally, this process was envisaged to occur by extension over a broad area, hence the term seafloor spreading[1]. With the benefit of high-resolution data, it has become evident that the basaltic igneous rocks making up the oceanic crust form within a few kilometres of the ridge axis. The crustal structure of the axial region is reasonably well understood from both observations and heat-flow theory: ridge axes act as zones of weakness concentrating further deformation because the deeper part of the crust is hot and weak. The deeper-mantle processes, however, are more difficult to observe and model, so two papers elsewhere in this issue[2,3] provide much-needed data.

In particular, it is not evident how the magma that will form the oceanic crust ascends from source regions in the mantle to the narrow axial zone (see figure). The difficulty arises as follows: pressure-release melting beneath ridges occurs as mantle material upwells adiabatically (the melting temperature decreases with de-



Cross-section of a ridge axis, shown with no vertical exaggeration. Features studied by Blackman and colleagues are shown on the right: the flow pattern aligns anisotropic crystals, and this can be detected from the different travel times for ray paths inside and outside the upwelling zone. Features studied by Ildefonse and colleagues are shown on the left: stress variations alternating with time produce dikes in perpendicular directions.

creasing depth). Crudely speaking, the rate of melt generation in a volume of rock is proportional to the vertical component of its velocity within the zone of extensive melting, which extends down to a few tens of kilometres depth. The width of upwelling is expected to be comparable to the depth. Therefore if melt ascends vertically to the surface, the zone of igneous activity is expected to be tens of kilometres wider than it is. Either the zone of upwelling is narrow or the magma must be able to move, in a lateral direction, into the axial region.

The papers by Blackman *et al.*[2] and Ildefonse *et al.*[3] both provide data directly related to this problem. The first discusses large long-term strains that produce fabric (ordered orientation of mineral grains) in the upwelling mantle beneath ridges. The second relates to the instantaneous stress field that determines the orientations of dikes in the mantle.

Blackman and colleagues[2] studied the travel times from distant earthquakes to an array of seismic stations on the sea floor near the Mid-Atlantic Ridge axis, and found that P-waves (primary waves, which are compressional and locally propagate upwards from great depths) arrived early at the ridge axis. This observation cannot be explained by the thermal structure of the ridge axis because the hot rocks have lower seismic velocities. Instead the authors propose that the mantle in the upwelling zone is anisotropic. This hypothesis is based on the well known observation that the oceanic lithosphere is anisotropic with the seismically fast direction horizontal and perpendicular to the ridge axis. Basically, olivine crystals (the chief mineral in the upper mantle) are anisotropic. Their fast direction is aligned in the transport direction within the plane of shear when the rock becomes rigid at the base of the lithosphere. The transport direction within the plane of shear is nearly vertical in the upwelling rock. The width of the vertical-fast anisotropic region — a few tens of kilometres — thus establishes that the upwelling region is much wider than the zone of igneous activity.

Ildefonse *et al.*[3] take a different tack, studying rocks in Oman that were once oceanic mantle. They have discovered an isolated upwelling region that was not subsequently deformed by horizontal shear at the base of the lithosphere and have deduced the orientations that the dikes (tabular, igneous intrusions) had in the upwelling region. Vertical dikes in the plane of the ridge axis, horizontal sills and some vertical 'A–C' dikes perpendicular to the ridge axis were mapped. These intrusions appear to have formed a network of molten channels that may have helped carry magma to the ridge axis. However, it is still not conclusive that such dikes (rather than porous flow through the partially molten mantle) are the dominant path that carries magma to the axial zone.

The mechanics of how several orientations of dikes coexist are not fully evident. Dikes are expected to intrude perpendicular to the axis of least compressive stress. This axis is roughly horizontal and perpendicular to the ridge at shallow depths where the lithosphere is being stretched. Ildefonse *et al.* propose two mechanisms that may act together to produce the horizontal sills and the A–C dikes. The first involves the excess pressure of the magma itself. This pressure often exceeds lithostatic pressure in the Earth — magma can rise to produce high volcanoes. Intrusion of a thick dike makes the stress more compressive in the direction perpendicular to the dike and may eventually cause the least compressive stress to lie along another axis, leading to a second set of dikes perpendicular to the first. Their second mechanism involves transient failure of the lithospheric lid either by crustal faulting or crustal dike intrusion which increases horizontal tension in the underlying mantle. The 'normal' asthenospheric stress field associated with mantle flow is present at other times. Ildefonse *et al.* call the alternate inflation of dikes in different orientations 'breathing'.

One aspect of this second explanation constrains the geometry of upwelling. For the typical least compressive stress to be vertical, either the strain rate in the vertical direction $\dot{\varepsilon}_z = v_z/z$ must be positive or the upwelling velocity must increase with shallower depths. The total volume flux per length of ridge axis of mantle material upwelling at a given depth clearly decreases upwards, reaching zero at the base of the crust. For upwelling velocity to increase upwards, the width of the upwelling region must decrease faster than the total volume flux.

Until now, models of the upwelling beneath ridge axes have by necessity been mixtures of elegant physics, based on simple assumptions, and qualitative observations of exposed mantle rocks. Blackman and colleagues' seismic study now indicates that the upwelling region can be sensed remotely, and the field study by Ildefonse and co-workers indicates that careful observations of dikes and fabric can constrain both transient and long-term processes in the upwelling. Both studies doubtless hold further implications, yet to be fully explored, for how ridges work. □

Norman H. Sleep is in the Department of Geophysics, Stanford University, Stanford, California 94305, USA.

1. Dietz, R. S. *Nature* **190**, 854–857 (1961).
2. Blackman, D. K., Orcutt, J. A., Forsyth, D. W. & Kendall, J.-M. *Nature* **366**, 675–677 (1993).
3. Ildefonse, B., Nicolas, A. & Boudier, F. *Nature* **366**, 673–675 (1993).

DAEDALUS

Home brewing

ALCOHOL is a treacherous friend. Many people value its mood-expanding qualities; but the heavy drinker risks serious medical trouble. Curiously enough, even the strictest teetotaller is exposed to alcohol, for it is naturally generated in the mammalian gut. The main culprit is the gut bacterium *Sarcina ventriculi*, which releases up to 30 g of alcohol a day into our intestines — equivalent to three 'standard drinks' such as glasses of wine or half pints of beer. We don't notice, because the liver mops up this small steady flow very efficiently. Nonetheless, we are all imperceptibly drunk all the time.

Dedicated alcoholics will be indignant at this wasted resource: for there is no point whatever in being imperceptibly drunk. Furthermore, the recommended safe daily alcohol intake for men is 30 g a day (women, with their less efficient livers, should keep below 20 g a day). Pooling internal and external sources, therefore, men can safely handle 60 g a day, and women 50 g. So, says Daedalus, if our internal source could be eliminated, men could safely drink six standard drinks a day, and women five drinks: a much more cheerful prospect. For by taking this dosage in a few large slugs rather than wasting it as a dribble, it could all be used to establish a period of worthwhile intoxication.

So DREADCO's biochemists are seeking a selective antibiotic to eliminate *Sarcina ventriculi* and other internal brewers from our guts. Already pilot compounds are being tested on experimental animals. They are making kittens noticeably less kittenish than usual, and guinea pigs less gormless — Daedalus suspects that these creatures are usually a bit drunk on their natural internal fermentations. Once perfected, 'Soberinsides' will be widely welcomed. Alcohol lovers will be able to double their consumption before hitting the medically prudent limits. Breweries and wine merchants will be able to sell twice as much before being accused of threatening the health of the community. Best of all, Soberinsides will bring new sharpness to our sober mental processes. For by eliminating the steady mental sabotage of internal fermentation, it will clear away the slight unnoticed alcoholic haze which now clouds our minds from birth to death. It will give all of us a calm clarity of thought, a keenness of insight and acuity of inference, equivalent perhaps to 10 or 20 extra points of IQ. Soberinsides will expand our mental life two ways: both reducing the perils of being really drunk, and allowing us for the first time to be really sober.

David Jones

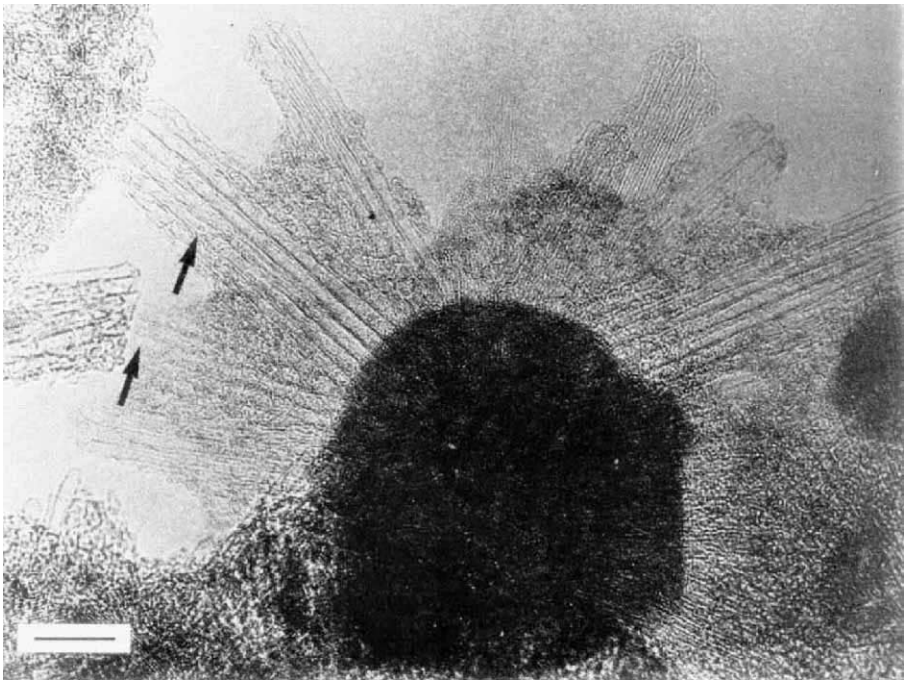
Radial single-layer nanotubes

SIR — Several novel forms of carbon have been discovered since the mid-1980s, including fullerenes[1], nanometre-sized nested graphitic tubes and polyhedra[2], and single-layer carbon nanotubes[3,4]. These single-layer tubes have so far been observed only in the primary soots from arc-discharge experiments where the anodes were packed with the ferromagnetic elements Fe (ref. 3), Co (ref. 4) and Ni (which was observed by our group). Here we report the discovery of clusters of single-layer tubes arranged around nuclei of Gd_xC_y in patterns similar to sea-urchin spines. They are present in large quantities in the primary soots from arc-discharge experiments using either Gd_2O_3 or pure metal Gd in the anode.

The soot samples described here were prepared in a manner similar to that described earlier[5]. The graphite anode was drilled out and packed with either Gd_2O_3 or Gd such that the Gd/C ratio was between 3 and 6 atom %. The cathode is a 12.7-mm diameter graphite rod, and the anode a 8-mm diameter graphite rod with a 3.2-mm bore drilled out in the centre. The arc gap was typically 3–6 mm, and the current and pressure were 85 amp and 500 torr (Gd_2O_3) or 75 amp and 1,000 torr (pure Gd). The primary soot obtained was dispersed ultrasonically in ethanol and placed on holey carbon-coated Cu grids for transmission electron microscopy.

The soot morphology can be broadly classified into two types — pure amorphous carbon particles; and amorphous carbon regions containing substantial quantities of high electron contrast metal nuclei. Energy dispersive spectra confirm that these nuclei are Gd compounds. Careful observation of the regions containing the Gd_xC_y particles shows the prevalence of a 'sea-urchin' type morphology, with the Gd_xC_y particles forming the nuclei of these radial structures.

The figure shows a high resolution image of one of the 'sea urchin' particles, clearly illustrating the bundles of rather short (<75 nm) single-layer tubes radiating out from a Gd_xC_y core, which appears to be amorphous. Some of the bundles also appear to contain what look like multi-layered tubes, as indicated by arrows in the figure. We have also observed similar structures (not shown) around single crystal nuclei, which are partially encapsulated with turbostratic carbon layers. Analysis of the *d*-spacings of the crystalline nuclei prove that the single crystal cores are α-GdC_2. Note that the bundles of single layer tubes are not present on those facets where layers of turbostratic carbon cover the α-GdC_2 crystals (data not shown).



High-resolution transmission electron microscope image of bundles of short predominantly single-layer carbon tubes encapsulating an amorphous core of Gd_xC_y. Scale bar, 10 nm. Arrows, possible multi-wall tubes. The microscope is a Phillips CM20 ultra-high resolution TEM.

Observation of 'hybrid' encapsulates, which have GdC_2 single-crystal cores and are partially coated with single-walled nanotubes, and in other regions by incomplete polyhedral shells, suggests that the thermal history of the encapsulate may determine its final form. Complete encapsulation by (nested) polyhedral shells has never been observed by us in primary soot samples, but rather only in the 'boule' growth on the cathode, such as for α-LaC_2 (ref 5) and α-GdC_2 (ref 6). The 'sea urchin' encapsulates reported here demonstrate that dramatically different morphologies can result even though the encapsulate is the same (in this case, single crystal α-GdC_2). This opens the possibility of studying the encapsulates which are formed in the primary soot.

We have also observed that it is possible to significantly enhance the fraction of these novel types of carbon coated encapsulates in the primary soots by oxidizing the soot to about 50% (carbon) mass loss, which shows that the bundles of single-layer tubes are substantially more resistant to oxidation than amorphous carbon.

We have prepared various primary soots produced with other rare earth oxides. We find that Nd is the only other candidate producing the structure described here, and that La, Ce, Pr, Sm, Sc, and Er do not. Further experiments are needed to understand these differences.

Shekhar Subramoney
DuPont Company, Experimental Station, PO Box 80228, Wilmington, Delaware 19880-0228, USA
Rodney S. Ruoff
Donald C. Lorents
Ripudman Malhotra
Molecular Physics Laboratory, SRI International, Menlo Park, California 94025, USA

1. Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. *Nature* **318**, 162–163 (1985).
2. Iijima, S. *Nature* **354**, 56–58 (1991).
3. Iijima, S. & Ichihashi, T. *Nature* **363**, 603–605 (1993).
4. Bethune, D. S. *et al. Nature* **363**, 605–607 (1993).
5. Ruoff, R. S., Lorents, D. C., Chan, B., Malhotra, R. & Subramoney, S. *Science* **259**, 346–348 (1993).
6. Subramoney, S. *et al. Carbon* (in the press).

Waxy tea

SIR — Major components of the scum on tea[1] are high-melting-point lipids, epidermal waxes which are standard waterproofing equipment on leaves of land plants such as *Camellia* (*Thea*). Boiling water takes them off, and they float to the surface in tea cups, where they solidify as the surface cools. After that, disturbances (such as insertion of a spoon) crack the thin surface layer into floes, easily visible by obliquely reflected light. As the tea is drunk, the waxy layers stick to the sides of the cup. They stain with tea phenolics, which go brown on oxidation. Milk components, casein and butter-fat, thicken them. One can liquefy, saponify and solubilize them with very hot water and detergent. (I do this every day.)

Ralph A. Lewin
Scripps Institution of Oceanography, La Jolla, California 92093, USA

1. Spiro, M. & Jaganyi, D. *Nature* **364**, 581 (1993).

Plastids better red than dead

SIR — Beyond the end of the genetic rainbow of plastids discussed in News and Views by Palmer[1] lies a dwarf plastid-like genome he failed on that occasion to mention — that of the malaria parasite (but see ref. 2). Palmer's description of the reduced genome of the parasitic plant *Epifagus* as "nearly extinct" (its genetic content is tiny by comparison with ordinary plastids), might lead some to assume that the malarial plastid remnant, which is only half its size again, is equally moribund or even dead. But for various reasons I believe it is still vital.

Since Palmer's News and Views was published, we have obtained further evidence[3] supporting the plastid ancestry of the malarial 35-kilobase circular DNA. The DNA encodes an open reading frame (ORF470) whose predicted peptide product is about 50% identical with a plastid ORF so far reported only in the red algae *Antithamnion spp.*[4], *Cyanidinium caldarum* (K. Zetsche, personal communication) and *Porphyra purpureum* (M. Reith, personal communication). The high level of conservation strongly suggests that ORF470 is a functional gene retained in the highly reduced malarial plastid remnant — its presence thus would fall in line with Palmer's rationale for retention of plastid genomes in non-photosynthetic plants[5]. This likelihood is strengthened by our finding that transcripts of ORF470 are readily detected in total malarial RNA, whereas transcripts of other ORFs encoded by the malarial plastid-like DNA, those encoding elements of the transcriptional machinery — ribosomal proteins and so on — are much rarer.

A second notable feature of the plastid-like DNA of the malaria parasite (still incompletely sequenced), is the retention, unlike *Epifagus*, of at least some elements of the plastid *rpo* operon[6,7]. The malarial *rpoB* gene has undergone considerable evolutionary 'drift' compared with extant chloroplast counterparts, yet it is both complete and expressed at a low level.

Finally, it is striking that a homologue of the 35-kb plastid DNA remnant of *Plasmodium* (malaria) has been retained in other parasites in the ancient phylum Apicomplexa, such as *Toxoplasma* and *Eimeria*, which are now phylogenetically distant to *Plasmodium*[8]. The degree of conservation of this plastid remnant remains largely to be discovered, but in the three genera mentioned above, the plastid-like DNA has a highly characteristic inverted ribosomal RNA repeat that can form a large cruciform structure containing spatially conserved sequences[9].

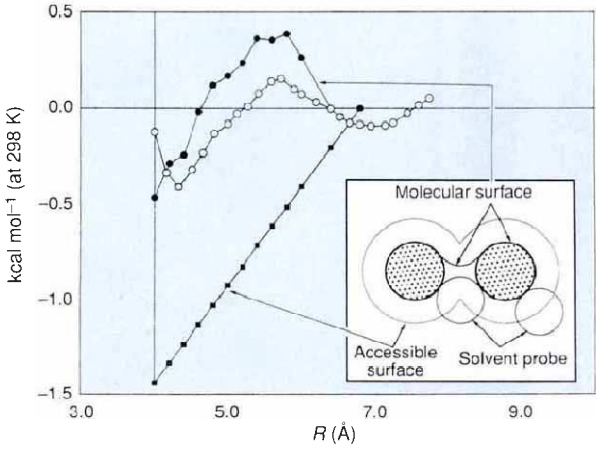
I suggest that whatever the case for the "nearly extinct" *Epifagus* molecule, the malarial plastid DNA remnant, which we have linked phylogenetically to the non-green end of the currently known spectrum of plastid DNAs, is not dead and that its function is succinct rather than extinct.

Iain Wilson
National Institute for Medical Research, Mill Hill, London NW7 1AA, UK

1. Palmer, J. D. *Nature* **364**, 762–763 (1993).
2. Palmer, J. D. *Curr. Biol.* **2**, 318–320 (1992).
3. Williamson, D. H. *et al. Molec. gen. Genet.* (in the press).
4. Kostrzewa, M. & Zetsche, K. *J. molec. Biol.* **227**, 961–970 (1992).
5. Wolfe, K. H., Morden, C. W. & Palmer, J. D. *Proc. natn. Acad. Sci. U.S.A.* **89**, 10648–10652 (1992).
6. Gardner, M. J., Williamson, D. H. & Wilson, R. J. M. *Molec. biochem. Parasit.* **44**, 115–124 (1991).
7. Howe, C. J. *J. theor. Biol.* **158**, 199–205 (1992).
8. Barta, J. R., Jenkins, M. C. & Danforth, H. D. *Molec. biol. Evol.* **8**, 345–355 (1991).
9. Wilson, I., Gardner, M., Rangachari, K. & Williamson, D. in *Toxoplasmosis* (ed. Smith, J. E.) NATO: ASI Ser. 51–62 (Springer, Berlin, 1993).

Protein surface area defined

SIR — Weber[1] in scientific correspondence suggested that the entropy of systems as diverse as proteins in solution and black holes are dependent on surface area. Langmuir[2] in 1925 observed that molecular surface area is related to the free energy of processes in solution. We believe that the definition of the surface area is critical in describing the energetics of association in solution. Accessible surface area (see figure) is widely used in protein modelling to describe the entropy-driven hydrophobic effect, but the approach fails to represent even the simpler manifestations of hydrophobicity[3].



Schematic of two methane molecules in close contact, representing the solvent accessible (ASA) and molecular surface areas (MSA) as trace out by a solvent probe. Potential of mean force for the dimerization of two methane molecules in water at 298 K, as represented by spheres of radius 2 and 1.4 Å, respectively. ■, Calculated using ASA with a surface tension of 24 cal mol^{-1} $Å^{-2}$; ●, using MSA with a surface tension of 102 cal mol^{-1} $Å^{-2}$. For comparison, ○ was taken from a Monte Carlo statistical mechanics simulation of the dimerization of methane[6].

Interactions in the condensed phase are characterized by an energy barrier for solutes separated by less than a single solvent molecule. Consider the dimerization of methane in water (figure). The potential of mean force generated by a linear relationship between accessible surface area and free energy is in disagreement with theoretical calculations and statistical mechanics simulations. Indeed, the force derived from the accessible surface area is attractive at close separation, because the surface must always be greater for two separated molecules than when they are in van der Waals contact.

In contrast, molecular surface area (see figure) reproduces the energy barrier. We have argued that for a given area and solvent, surface tension described by molecular surface area is shape-independent[4], in contrast to surface tension relating to accessible surface area[5]. The apparent differences in measures of microscopic and macroscopic surface tension can be resolved using molecular surface area, without the need to invoke a marked curvature dependence of surface tension.

This proposal has important implications for modelling surface-area dependence of protein–protein interactions , as calculations show that accessible surface area is not simply related to molecular surface area by a constant of proportionality.

Richard M. Jackson
Michael J. E. Sternberg
Imperial Cancer Research Fund, PO Box 123, London WC2 3PX, UK

1. Weber, G. *Nature* **365**, 792 (1993).
2. Langmuir, I. *3rd Colloid Symp. Monogr.* p.48 (Chemical Catalog Co., New York, 1925).
3. Wood, R. H. & Thompson, P. T. *Proc. natn. Acad. Sci. U.S.A.* **87**, 946 (1990).
4. Jackson, R. M. & Sternberg, M. J. E. *Protein Engineering* (in the press).
5. Sharp, K. A., Nicholls, A., Fine, R. & Honig, B. *Science* **252**, 106 (1991).
6. Pangali, C., Roa, M. & Berne, B. *J. chem. Phys.* **71**, 2975 (1979).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of fewer than 500 words and five references.

"Labour of the decypherer"

R. V. Jones

Codebreakers: The Inside Story of Bletchley Park. Edited by F. H. Hinsley and Alan Stripp. *Oxford University Press: 1993. Pp. 321. £17.95, $25.*

"THE Art of Decyphering", wrote Francis Bacon, "is a Work of Labour and Ingenuity, . . . if the Cyphers in use were good and trusty, several of them would absolutely elude the Labour of the Decypherer; . . . But through the ignorance and unskilfulness of Secretaries and Clerks, in the Courts of Princes, the most important Affairs are generally committed to weak and treacherous Cyphers."

Codebreaking has long been a British tradition, much enhanced in the First World War by the work of Admiralty Room OB40 (founded by the physicist-engineer Aldred Ewing) which culminated in the decoding of the Zimmermann telegram which brought the United States into the War in 1917.

The best work is often done in temporary huts rather than in pretentious buildings, and this was certainly true of Bletchley. The estate of Bletchley Park ('Station X' or 'B.P.' as it came to be called) had been selected as the emergency headquarters for MI6, should that organization have to be evacuated from Whitehall. As in 1939 the Government Codes and Cyphers School (the forerunner of GCHQ) was administered through MI6, Bletchley became the wartime home of the cryptographers.

Among the types of cipher mentioned by Bacon were wheel-ciphers, and in the Second World War the German services relied on a sophisticated development along these lines, the Enigma machine. The employment of such machines tended to dismay older codebreakers, who thought that their art would henceforth be useless; but fortunately there were younger men who believed that there might still be hope. Among these, first credit must go to the small Polish team who, as early as 1932 under Martin Rejeweski, 'broke' an Enigma message. They went on to build replicas of the Enigma machine, which in July 1939 they presented to the British and French cryptographers. The former had already made progress towards deciphering Enigma, and took full advantage of the Polish gift, which included the secretly wired Enigma wheels. It was at first believed at Bletchley that the Poles had stolen the wheels from a German machine, but it transpired that they had succeeded in the very much more difficult feat of divining the wiring inside the wheels from a brilliant analysis of intercepted Enigma messages.

By May 1940 Enigma decrypts were beginning to emerge from Bletchley. The story of how these decrypts were used to assist British and American forces, above all in the Battle of the Atlantic, but also in the electronic war and in military and other operations, has been told in the official history of *British Intelligence in the Second World War*. But because of its intense secrecy, the work of the men and women at Bletchley and in its supporting organizations such as the signals interception stations received no public recognition for many years after the war, and many of those who made important contributions are no longer alive. This volume of personal recollections by some 30 of the survivors is therefore especially welcome.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

The Science Museum

Many variations — the Enigma cipher.

Those at Bletchley concerned with the Enigma ciphers used by the German air force and army worked in Hut 6, and German naval Enigma was tackled by Hut 8. The decrypts made by Hut 6 were passed to Hut 3, where they were emended and edited, and then forwarded — when appropriate — to the army and air staffs. Decrypts from Hut 8 passed to Hut 4, and thence to the Admiralty.

Survivors from each of those huts have contributed to *Codebreakers*, the greatest interest of course attaching to descriptions of the work in Huts 6 and 8, where the crucial steps made in the actual cryptography are outlined. Lest anyone should be tempted to underrate the feats achieved by the cryptographers, these so astonished their German counterparts that when *The Ultra Secret* was published in 1974 they at first refused to believe that it was true. Only when Gordon Welchman, the first head of Hut 6, gave details of its methods in *The Hut Six Story* in 1982 did a German cryptographer tell a former member of Hut 6 "Now we really do believe that you did it!".

Welchman was succeeded as head of Hut 6 by Stuart Milner-Barry, whose article throws much light on the technique of 'breaking'. This exploited both the arcane weaknesses of the Enigma machine itself (in particular, no letter could appear as itself in the enciphered text), and of the German operators who used it (such as not changing the wheel settings between messages, or using standard phrases in pro forma reports). Milner-Barry states that "[h]ad it not been for human error, compounded by a simple design quirk, the Enigma was intrinsically a perfectly secure machine", thus exemplifying Bacon's warning of the insecurity arising from "the ignorance and unskilfulness of Secretaries and Clerks".

Fantastically difficult as Enigma (especially the naval version with four, rather than three, wheels) was to break, Bletchley also tackled two further machines (which it named 'Fish') introduced by the Germans. Whereas Enigma signals were transmitted by Morse code, the new machines used teleprinter techniques employing the digital encoding of letters. The first, 'breaks' into Fish were made by Brigadier John Tiltman in the spring of 1942, and in the following year work proceeded much more speedily with the aid of Colossus, famous as a forerunner of the modern computer.

The work at Bletchley was so vast and so deep that no short review can do it justice. Besides Enigma and Fish, there was work on the simpler German ciphers (which besides being tactically valuable in itself sometimes gave clues to the content of Enigma messages) and on those of the Allies' other opponents, while the principal Japanese ciphers were tackled very successfully by the Americans. While the actual cryptography was the key activity, much back-up was needed to give the cryptographers the best briefs, and to make best use of the decrypts. The design of the machine aids to cryptography and the related electronics owed much to C. I. Wynn Williams, the inventor of the scale-of-two counter (originally for nuclear physics), and to Alan Turing and T. H. Flowers. Chess players and Cambridge geometricians such as Gordon Welchman were conspicuous among the codebreakers, while many scholars from arts faculties, such as Geoffrey Barraclough (history), F. L. Lucas (English literature) and

Frederick Norman (German) staffed Huts 3 and 4.

Biological scholarship too found a niche. Geoffrey Tandy, a keeper in the botany department of the Natural History Museum, had enlisted in the Royal Naval Volunteer Reserve. He declared that he was sent to Bletchley because when the services were scoured for possible recruits, the attention of Their Lordships of the Admiralty focused on him because he had stated his civilian occupation as 'cryptogamic botanist' — it was therefore considered that he would be adept at cryptography. Posted to the naval section at Bletchley, he quickly established himself as a 'character' with such epigrammatic observations as "The bigger the foot, the splasher!" and he built up a dossier of what he called 'linguistic equivalents' which listed the nearest terms in other languages to the new terms that were coming in such as 'radar' or 'schnorkel'. At the end of the war he had to plead his case before an Admiralty board for the continuation of his work in peacetime. After a dissertation on the importance of precise language and definitions he illustrated it by telling the admirals "For example, gentlemen, if I were to say *troglodites troglodites troglodites* every zoologist would know that I was referring to the common wren!"

Much more could have been written about Bletchley, with its atmosphere of academic liveliness, had the survivors been able to tell their stories earlier. Indeed, it is noteworthy how loyally they kept silent, even not telling their wives and families. Walter Eytan describes the shock to many of them when in 1974 *The Ultra Secret* was published without their receiving any advance notice.

Conditions of life and work at Bletchley, and its principal achievements, are faithfully sampled in *Codebreakers*, which is worth reading both for its historical interest and for the sidelights it throws on the problems encountered in the rapid assembly and organization of one of the greatest collections of talent that has ever occurred in Western civilization. There is still room left for a reflective analysis of these problems, before memories of Bletchley are laid to rest in accord with the title its staff gave to their Christmas Review in 1944 — 'R. I. B. P.' □

R. V. Jones is at 8 Queen's Terrace, Aberdeen AB1 1XL, Scotland, UK.

A galactic astronomer

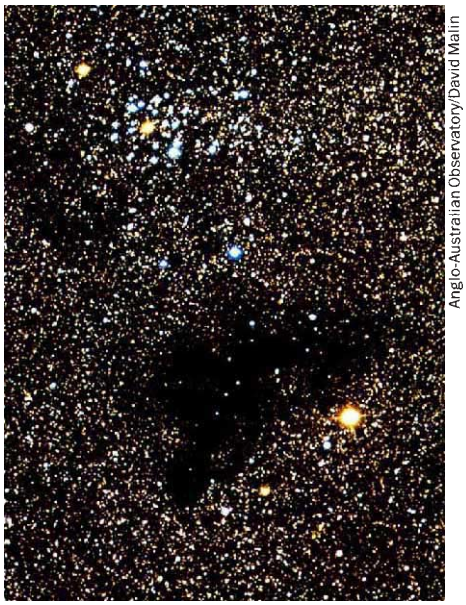
David W. Hughes

The Man Who Sold the Milky Way: A Biography of Bart Bok. By David H. Levy. *University of Arizona Press: 1993. Pp. 247. $35.00; £29.95.*

THERE are always two characters in any biography — the subject and the biographer. Why did Levy choose Bok? And in doing so, are the attributes that he found interesting the ones that we really want to know about?

The first question is easy to answer. Bok 'fired on six cylinders'. As such he was much more than just a first-class astronomical observer, an inspiring teacher and research leader, a competent and keen director of observatories, a popularizer of astronomy with a compelling writing style and an attention-grabbing lecture delivery, an allegedly left-of-centre scientific political activist and an avid opponent of pseudosciences such as astrology. He was all of these things at the same time. And Levy and Bok both lived in Tucson, Arizona.

Bartholomeus Jan Bok was born in Hoorn, on the western shore of Holland's Zuider Zee, on 28 April 1906. In 1924 he went to Leiden University to study mathematics, physics and astronomy, following this with graduate studies at the University of Groningen under the supervision of Pieter van Rhign. He became enthralled by the contemporary debate between Harlow Shapley and Jacobus C. Kapteyn as to whether the Sun was near the centre of the Milky Way or in its outskirts. Shapley was Bok's mentor and in 1929 gave him a job. Bok moved to the Harvard Observatory, first as the George Russell Agassiz Fellow and then as the Robert W. Wilson Teaching Fellow.



Anglo-Australian Observatory/David Malin

Birth place of stars — Bok globules, dark clouds against the stellar background.

Bok's study of the grand sweeps of swirling gases that surrounded the star Eta Carinae in the Milky Way won him his doctorate in 1932. By the mid-1930s he was deeply involved in all aspects of Harvard astronomy — teaching, observing, research and observatory politics. In 1946 he became associate director of the observatory.

Bok was always interested in the dark absorbing material in the arms of the Milky Way. Around 1947, together with Edith Reilly, a technical assistant, he began studying the dark nebulae that had been catalogued by E. E. Barnard. The very small, round and unusually dense ones especially caught their interest. Bok was a great lover of observing and started to photograph these with Harvard's Jewitt Schmidt telescope. He soon became convinced that the nebulae marked the birth place of stars, and they became known as Bok Globules.

Bok was a founding member of both the American Association of Scientific Workers and the International Council of Scientific Union's committee on science and its social relations. He wanted to increase the public awareness of science and to encourage astronomers to communicate with members of the public by writing, lecturing and broadcasting. Bok worked diligently to add the 'S' for Science to the newly founded United Nations Educational and Cultural Organization. As a result of his efforts, Bok suffered considerable harassment from Senator Joseph McCarthy.

It was widely expected that Bok would follow Shapley as director of the Harvard Observatory but it was not to be. Bok became disenchanted. Even though Harvard offered him the chair in their School of Education he turned it down in favour of a job at a quarter of the salary. He told the dean "I am an old Milky Way boy, and an old Milky Way boy I want to stay".

Bok moved to Australia in 1957 to become professor of astronomy at the National University and director of the Mount Stromlo Observatory. He did much to promote the founding of the Anglo-Australian Observatory. In 1966 he decided to return to the United States as head of the department of astronomy and director of the Steward Observatory at the University of Arizona. He retired in 1970 at the age of 64 and died in August 1983.

David Levy interviewed Bok more than 50 times during Bok's last two years of life, and has written a most enjoyable and readable biography. But it lacks bite. I kept feeling that Levy was acting as Bok's amanuensis and instead of providing the reader with a 'warts and all' view, everything is somewhat sanitized. As a typical example, Bok was an inveterate traveller and Levy quotes him as declaring "I want to pee in all the great rivers of the world".

Surely Levy must have asked him why. Other travellers are not like that. Now I'll never know the answer.

Bok was clearly a man of great character but we are not told what drove him. Levy concentrates on his life more than his science. I would have liked to know more about Bok's writings, and why he wrote what he wrote, and how it influenced the advancement of astronomy. The reader is told indirectly a great deal about what Bok thought of Bok, and what Levy thought of Bok, but I would have liked to have learnt more about what other astronomers thought of Bok. □

David W. Hughes is in the Department of Physics, University of Sheffield, Sheffield S3 7RH, UK.

Propaganda of symbiogenesis

T. Cavalier-Smith

Concepts of Symbiogenesis: A Historical and Critical Study of the Research of Russian Botanists. By L. N. Khakhina. Edited by L. Margulis and M. McMenamin. Translated by S. Merkel and R. Coalson. *Yale University Press: 1992. Pp. 177. $37, £22.50.*

THE idea that new organisms could arise as chimaeras by the fusion of separate organisms goes back at least to the ancient Greeks. It began to be taken seriously by scientists only with the discovery by the Russian Andrie S. Famintsin (1835–1918) and the German S. Schwendener (1869) that lichens are composites of a fungus and an alga. This led Heinrich A. de Bary to adopt in 1879 the word 'symbiosis' for such relationships. Another Russian, K. S. Mereschkowsky (1851–1921), coined the term 'symbiogenesis' for the idea that symbiosis could lead to the permanent fusion of two distantly related organisms. Famintsin and Mereschkowsky both believed that symbiogenesis was an important and widespread evolutionary mechanism, as did a third Russian, Boris M. Kozo-Polyansky (1890–1957), and the American Ivan E. Wallin (1883–1969), who even thought it was the basis for all speciation in eukaryotes. The ideas of these scientists are the focus of this book.

The evolutionary importance of symbiogenesis in rare instances is now well established, with the acceptance of Mereschkowsky's theory of the symbiotic origin of chloroplasts from cyanobacteria (then called blue-green algae); of the theory of the symbiotic origin of mitochondria from bacteria, developed independently by Wallin and Kozo-Polyansky; and of the theory of the symbiotic origin of the kingdom Chromista (cryptomonads, chlorarachniophytes, haptophytes and heterokonts — for example, brown algae, diatoms, chrysomonads) — from two separate eukaryotes. The symbiotic origin of peroxisomes and glycosomes is, however, still only hypothetical.

This translation of L. N. Khakhina's historical work published in Russian in 1979 is at first sight most welcome particularly as US researchers often wrongly attribute Mereschkowsky's theory of the multiple symbiotic origin of chloroplasts to Raven and Margulis. In the past dozen years, three US researchers have independently proposed theories of the symbiotic origin of the nucleus as if they were novelties, not realizing that this erroneous idea goes back at least to Boveri in 1904, and probably well into the nineteenth century; it was espoused particularly strongly by Mereschkowsky and Kozo-Polyansky. In their lengthy introduction, the editors attribute this ignorance, overcharitably, to ignorance of Russian in the United States. But Mereschkowsky's seminal 1905 and 1910 papers were in German and his long review in 1920 was in French. All have been widely cited in English language publications.

Unfortunately, the book — especially its introduction — is riddled with propaganda and error, and the longest chapter on 'contemporary concepts' is 15 years out of date. For a start, Alexander Vacinich in his foreword badly misrepresents Mereschkowsky's views by stating that he thought that all cells are symbiotic combinations. He did not: he excluded both bacterial and fungal cells from the idea of a dual origin.

Margulis has vigorously promoted two major confusions that permeate the book. The first confuses the symbiotic origin of mitochondria (true) with the symbiotic origin of the eukaryotic cell (probably false). Khakhina's book is neither as critical nor as historically accurate as it should be. She often refers to "the origin of the cell by symbiosis" when she means either the origin of nuclei or the origin of chloroplasts.

In her introduction, Margulis states that Khakhina's main point is that "the early symbiogeneticists believed Darwinism and natural selection to be useless or irrelevant to the concept of evolutionary change". On the contrary, Khakhina emphasizes that Famintsin thought that "natural selection provided a singularly plausible and completely satisfactory explanation of the causes of the formation of adaptations", that Kozo-Polyansky considered "natural selection as the primary mechanism of the development of the organic world", while Mereschkowsky "thought it necessary to correlate it [symbiogenesis] with the teaching of Darwin. This he did with the utmost brevity." Although she states that Mereschkowsky thought earlier theories (those of Darwin, Haeckel and Naegeli) were "unsuccessful", "outmoded" and "did not agree with" symbiogenesis, she does not make his exact criticisms clear.

Newton Wonder — one of over 2,000 apple varieties surveyed in *The Book of Apples* by Joan Morgan and Alison Richards with paintings by Elisabeth Dowle. Ebury Press in association with the Brogdale Horticultural Trust, £19.99.

Margulis's second major confusion has been to muddle the pervasive role in evolution of symbiosis as a factor influencing the mutual coadaptation of symbiotic partners with the extremely rare permanent merging of such partners to form a single organism — that is, symbiogenesis in Mereschkowsky's original sense. Unfortunately, Khakhina makes the same mistake and uses the term symbiogenesis for both roles. Her claims that "symbiosis is emerging . . . in the world of prokaryotes as a major mechanism of evolution" is totally misleading. There is not a single known example of cellular endosymbiosis in prokaryotes, and no good reason to think that two different prokaryotes have ever merged together to form a single new organism.

Famintsin did not think, as we do now (and as Khakhina sometimes implies), that chloroplasts were organelles that had evolved from symbionts, but that they are symbionts: he spent years vainly trying to cultivate them, as did Wallin for mitochondria. The idea that chloroplasts and mitochondria were permanent organelles that never arose *de novo*

evolved only gradually: though some people thought this a century ago, it was established only in the 1960s.

Khakhina is loth to make historical judgements, writing that Elenkin's view that Famintsin conceived symbiogenesis long before his discovery of the dual nature of lichens, and P. Borodin's totally contradictory view that the lichen discovery led to the idea of symbiogenesis, both "seem just": clearly at least one must be wrong. A reluctance to take sides or to be critical of dogma is perhaps understandable for someone writing in the Soviet Union of the 1970s. Equally understandable is her praise of symbiogenesis as approaching "a truly dialectical understanding of the factors of organic evolution" and her nationalistic conclusion that "Russian scientists played a leading role in developing the concept of symbiogenesis, particularly in the early stages"; this exaggeration appears plausible only because the book largely ignores non-Russians, except for Margulis, whose contributions are overstated.

One day, there will be a decent history of research on the origin of cell organelles by symbiogenesis. This one, I am afraid, is too distorted by propaganda. □

T. Cavalier-Smith is in the Department of Botany, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4.

La dolce vita?

Giovanni F. Bignami

L'università dei Tre Tradimenti. By Raffaele Simone. *Laterza: 1993. Pp. 151. L13,000.*

FULL professors in Italian universities, some of them unfondly known as *baroni*, all lust for power. Some have it but none of them really control the powers that rule over their own university and which could (if properly directed) make it work. But how does one become a full professor in Italy? Easy, just follow the few written and many unwritten rules clearly given in the chapter "How to get in". You'll learn about the laws of "primordial affiliation", of "exchange of academic gifts", of "rewarded loyalty" and more. And if you think that these are self-explanatory, beware: Italian academia has more facets than you or I can hope to fathom. Even in his courageous and brilliant pamphlet, Simone, lest he bore the reader, shies away from describing precisely how to get a chair in an Italian university. You actually get it by winning a *concorso*, a national competition judged by ministerially appointed committees where the needs of individual universities are diluted if not lost to academic bargaining at best.

Some time ago, D. Burr explained the process (*Nature* **357**, 273; 1992); he used a whole page and still only covered part of the subject. He also gave up on translating the Italian word *concorso*, the heart of the professorship procedure. (It is actually best grasped from its obvious Latin origin: *cum correre*, to run together ....) That exposé was indeed necessary in the wake of a flurry of letters, obscure, I suspect, to the average *Nature* reader, from Italian academics unhappy about the outcome of the last round of *concorsi*. By necessity, most of those letters were grouped together under 'academic promotion'. To an Italian ear, however, 'promotion' sounds a totally inadequate rendering for the radical change in official and social status that comes with winning a *concorso*.

It will be a challenging, and very useful, task to translate into English this book by Simone, himself a linguist. At the moment, you not only need to be proficient in Italian, but also be conversant with classical Greek, for example to appreciate the concept of "*nòstos*", one of the plagues, according to Simone, of our universities. Heroes in classical mythology had to spend a large part of their life wandering in desolate and hostile places doing, by definition, heroic deeds. Eventually they felt they had earned their right to and started longing for, their *nòstos*, a triumphal and touching homecoming. Not that Italian university professors have much in common with Greek heroes but, after winning their *concorso* and serving out their prescribed time in a small university, they all want to come back to their parent institution, usually one of the big ten, out of the 60 or so in the country. But positions in big universities are few and coveted, so here is where the 'old-boy' network comes in handy, or better still, membership of the freemasons or Opus Dei. (Simone, however, makes a point in his opening quotation of Karl Kraus, that "it is not always appropriate to name names ....") J. LaPalombara, a student of things Italian and a personal friend of a former prime minister (also, of course, a full professor), calls the system one "of strict cooptation" (*Democracy Italian Style*, Yale University Press; 1987).

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

The Ronald Grant Archive

The 'old-boy network' can come in handy.

The professors' *nòstos* is bad for the students, towards whom the most important of the three treasons is directed. It may, by contrast, be negligible, considering Simone's postulate on indifference towards students: "no one cares about students in Italian universities". A concept that does no justice to many a hard-working and available colleague, but which is, on average, sadly true, and especially so in the small, "provincial" universities. And it is not just the professors who are at fault. The whole ministerial machinery wastes human and practical resources in local mini-campuses, underdeveloped and underattended, frequently only created for local, quasi-political interests. This blatant waste of limited state resources is the second of the three treasons, not always easy to appreciate for outsiders, and one which escaped the poignant analysis, now a few years old, of B. R. Clark (*Academic Power in Italy*, Chicago University Press; 1977).

Now for the third treason: that towards research, by law one of the *raisons d'etre* of the Italian university system. It may, in selected cases, be of some relevance, but it is more frequently "irrelevant" and, paradoxically, is by and large impossible, owing to lack of funds and the abundance of red tape. Simone, a humanist bordering on science, draws here a fine line between humanities and science. In the former, irrelevant research is more frequent than in the latter, and not just because of the keen international competitiveness of science. Physicists in particular and scientists in general, we learn, are "more attached to their trade", for reasons which, Simone says, defy explanation. (How about "because it's there"?)

A delightful and important book, not about corruption but about reality in the Italian university system. It is easy to predict that it will become a bestseller at least in Italy though, of course, not among our students, the logical target for the book. Rather among professors, keenly bent on discovering the various treacheries of their own colleagues. □

Giovanni F. Bignami is in the Istituto per Ricerche in Fisica Cosmica e Tecnologie Relative, Consiglio Nazionale delle Ricerche, Vai E. Bassini 15, 20133 Milano, Italy.

Proteins regulating Ras and its relatives

Mark S. Boguski & Frank McCormick

GTPases of the Ras superfamily regulate many aspects of cell growth, differentiation and action. Their functions depend on their ability to alternate between inactive and active forms, and on their cellular localization. Numerous proteins affecting the GTPase activity, nucleotide exchange rates and membrane localization of Ras superfamily members have now been identified. Many of these proteins are much larger and more complex than their targets, containing multiple domains capable of interacting with an intricate network of cellular enzymes and structures.

RAS proteins and their relatives play critical roles in the control of normal and transformed cell growth, and have therefore been among the most intensively studied proteins of the past decade. Over this time, however, it has become clear that small GTPases of this type in fact control an extraordinarily wide variety of cellular processes, serving to monitor and control much of the information flow in eukaryotic cells. In addition to their role in regulating many aspects of growth and differentiation, members of this superfamily play an important part in control of the cytoskeleton and are central to the regulation of traffic between various membrane-bound compartments in the cell. These proteins have properties that make them uniquely suitable for such tasks: not only do they function as binary switches, cycling between active and inactive states bound to GTP and GDP, respectively, but also this cycling can be influenced by three different classes of proteins (which switch the GTPase on, switch it off and protect it from switching, respectively). Each of these regulatory proteins can itself mediate a variety of effector functions, some of which may affect further GTPases involved in different regulatory pathways. As the intracellular location of the GTPase can be controlled by a fourth class of regulatory protein, affecting the regulators with which the GTPase can interact, these proteins have an exceptional degree of information 'fan-in' and 'fan-out', allowing them to integrate inputs from a wide range of cellular pathways and contribute to their control. In many cases, the regulators with which they interact to accomplish this are only beginning to be understood, but already some significant insights and generalizations are starting to emerge.

In 1980, two *ras* genes were known: the first, now known as v-H-*ras*, is the transforming gene of Harvey, BALB and Rasheed sarcoma viruses, and the second, v-K-*ras*, is the transforming gene of Kirsten sarcoma virus[1]. By 1990, 40–50 distinct Ras-like proteins controlling many different aspects of cell growth and behaviour had been identified in mammals as well as many others in fungi, flies, frogs and nematodes. It had also become clear that Ras proteins are members of an extended family of GTPases[2,3] which includes proteins involved in protein synthesis (the elongation and initiation factors) and signal transduction (heterotrimeric G proteins). Figure 1 shows a simplified representation of these proteins' switching cycle; details of each step have been described elsewhere[2]. The GTP-bound form is slowly converted to the GDP-bound form by the protein's intrinsic capacity to hydrolyse GTP, a process accelerated by GTPase-activating proteins (GAPs). Activation in turn involves the replacement of GDP with GTP, an event mediated by proteins known as guanine-nucleotide-exchange factors (GEFs; also known as guanine-nucleotide-releasing proteins (GNRPs) or guanine-nucleotide-dissociation stimulators (GDSs)), and inhibited by guanine-nucleotide-dissociation inhibitors (GDIs). The latter may also block the action of GAPs. The anchoring of GTPases to various cellular membranes is also critical for their proper function, and is regulated by prenyltransferases, among other enzymes.

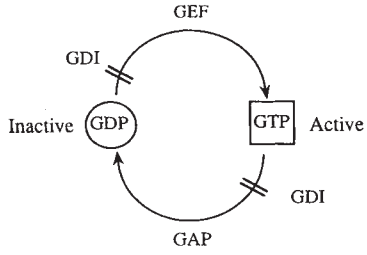


FIG. 1 Regulation of Ras-like GTPases by GEF, GDI and GAP proteins GEFs catalyse exchange of GDP for GTP: GAPs catalyse conversion of GTP-bound forms back to their inactive GDP states. GDI proteins for Rab and Rho affect nucleotide dissociation and GAP attack, but are thought to have other major roles in membrane localization and solubility.

Insofar as any protein may be considered 'simple', Ras and its relatives are rather small and uncomplicated compared with the majority of proteins with which they interact (Table 1). Typically, they contain about 190 residues, most of which are folded into a single GTPase domain[2]. In contrast, GAPs, GEFs and GDIs may be 10–15 times larger and are often modular structures containing several different types of functional domains. Here we will focus on domains and motifs that are conserved among particular subfamilies of *ras*-regulatory proteins.

The Ras superfamily of GTPases

The structures and functions of Ras-like GTPases have been reviewed elsewhere[1,5]. Based on their sequences, Ras-like proteins fall into three main families, each of whose members have similar functional characteristics.

Ras proteins. Mammalian cells express four 'true' Ras proteins (encoded by H-*ras*, N-*ras*, K-*rasA*, K-*rasB*), each of which can function as an oncogene following mutational activation. These proteins transmit signals from tyrosine kinases at the plasma membrane to a cascade of serine/threonine kinases, which delivers signals to the cell nucleus. Constitutive activation of this pathway contributes to malignant transformation.

The true Ras proteins are closely related to members of the Rap subfamily (Rap1A, 1B, 2A and 2B), as well as R-ras, RalA, RalB and TC21. Rap1A is also known as k-Rev 1 (ref. 3), a name stemming from its ability to reverse cellular transformation induced by K-*ras* and Smgp21 (see ref. 6 for guidelines on nomenclature). It is 53% identical to the H-*ras*-encoded p21, and may function by competing with Ras proteins for interaction with their effectors.

R-*ras* is 55% identical to H-*ras*[2] and has an effector binding domain identical to that of the true Ras proteins. It cannot itself transform cells, but, unlike other known Ras proteins, binds through its C terminus to Bcl-2, a protein that regulates apoptosis[7]. The functions of RalA, RalB and TC21 are unknown.

Rho/Rac proteins. Proteins of this subgroup (RhoA, RhoB, RhoC, Rac1, Rac2, CDC42Hs and TC10) are involved in organization of the cytoskeleton. Rac is required for membrane ruffling induced by growth factors (a result of actin reorganization at the plasma membrane) and the subsequent formation of actin stress fibres requires Rho[8,9]. In *Saccharomyces cerevisiae*, the CDC42 product controls cell polarity, another process in which actin is involved. In addition to these actin-related functions, Rac proteins are essential components of the NADPH oxidase system that generates superoxide in phagocytes[10].

Rab proteins. Members of this subgroup regulate membrane 'trafficking', that is, the transport of vesicles between different intracellular compartments[11]. To date, 24 different Rab proteins have been identified, each of which may be specific for one type of membrane or vesicle.

In addition to these three major families, further subgroups typified by Ran and Arf have recently been identified (Ran and TC4; Arf1–6). Ran proteins are abundant nuclear GTPases involved in aspects of mitosis[12]; Arf (ADP-ribosylation factor) proteins are necessary for efficient ADP-ribosylation of $G_{s\alpha}$ (the

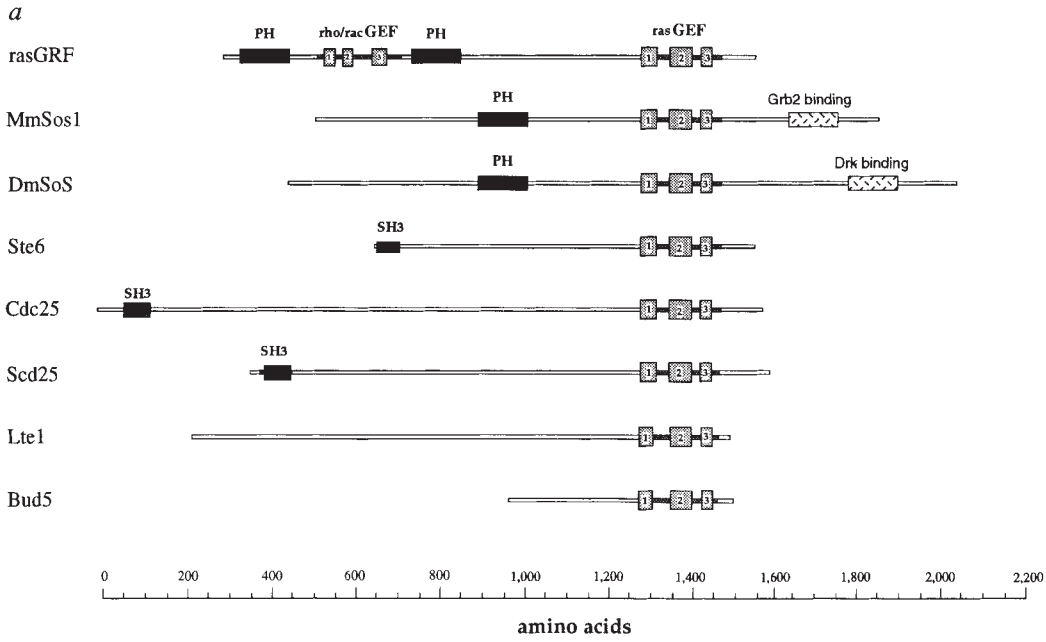


FIG. 2 Conserved domain in rasGEF and related proteins. Sequences were analysed with the multiple alignment tool kit[74] and MACAW program[125] as described previously[5,82]. Sources of the sequence data are indicated in Table 1. *a*, Schematic diagram of sequence relationships. Sequence diagrams are drawn to scale and aligned on the presumptive rasGEF domain. The grey rectangles labelled 1–3 represent the most highly conserved regions within the domain as detailed below. The locations of the sequences resembling Src-homology 3 domains are indicated for Ste6, Cdc25 and Sdc25. The three rectangles near the amino terminus of rasGRF represent the presumptive rhoGEF domain as detailed in Fig. 3. The pleckstrin homology (PH) domains in rasGRF, mouse Sos1 and *Drosophila* Sos are according to Musacchio *et al.*[45]. The hatched boxes labelled 'Grb2p-' or 'Drk binding', correspond to residues 1,148–1,304 and 1,342–1,550 in mouse Sos1 and *Drosophila* Sos, respectively. These domains contain four or five proline-rich repeats each[26,27]. *b*, Optimal alignments. Numbers along the left and right margins represent amino-acid residues encompassing the locations within the parent proteins of the presumptive rasGEF domains. Sequence conservation is not uniformly distributed throughout the domain but rather is concentrated in three discrete homology bloods or structurally conserved regions ('SCR') as shown in *a*. The *P* values associated with the various SCRs indicate the probabilities that the aligned regions are related by chance. (However, to eliminate the overestimation of statistical significance due to the substantial pairwise similarities of Cdc25/Sdc25 and Sos/mSos, the latter sequence from each pair was excluded from the calculations.) Consensus sequences are conservatively derived based on residue columns whose average alignment score (for all eight sequences) is at least 50% of the maximum possible score using the PAM120 matrix[123]. Dots in the consensus sequences represent somewhat more variable positions that nevertheless may be well conserved with respect to residue type, relative hydrophobicity, etc. A diagnostic signature for the rasGEF domain was derived from SCR3 and is represented in PROSITE[121] notation as: [GAP]-C-V-P-[FY]-×4-[LIVMFY]×-[DN]-[LIVM] The PROSITE Documentation entry number for this sequence motif is PDOC00594.

GTPase subunit of s-type heterotrimeric G proteins) by cholera toxin[2] and are thought to be involved in membrane vesicle fusion and transport[13–18].

Guanine-nucleotide-exchange factors (GEFs)

rasGEFs. Proteins that activate Ras proteins by exchanging bound GDP for free GTP are shown schematically in Fig. 2. In *S. cerevisiae*, *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Caenorhabditis elegans*, genetic loss of GEF function has similar effects to those induced by loss of the Ras proteins themselves, showing that GEFs are essential for Ras action. Loss of GEF function can be circumvented by mutations that constitutively activate the Ras proteins ('oncogenic' mutations) or, in some cases, through loss of GAP activity.

To facilitate nucleotide exchange, GEFs first associate with the GDP-bound form of the GTPase. GDP dissociates from this complex at an increased rate, leaving the GEF bound to the empty GTPase. GTP then binds immediately, prompting GEF dissociation and leaving the GTPase in the active form[3]. According to this scheme, a stable complex should exist between GEF and GTPase in the absence of nucleotide. Ras has indeed been shown to form such a complex with Cdc25 (ref. 19), as has RCC1 with Ran (see below). In both cases, GTP and GDP were equally effective at disrupting the complex, suggesting that GEFs can associate with both the GTP- and the GDP-bound forms of GTPase. GEFs have indeed been shown to recognize both the GDP- and GTP-bound forms of Ras[19,20], at least *in vitro*, although they seem to recognize the GDP-bound form more frequently *in vivo*[21], possibly because of other cellular factors.

Several mutants of Ras act in a dominant negative manner to block normal Ras activation (such as N17 H-*ras*[3]); these have reduced affinity for GTP and are probably defective in the final step of the exchange process (displacement of GEF by GTP), sequestering GEFs into dead-end complexes[5,19]. These mutants have been used to remove GEF activity from cells so that activation of endogenous Ras proteins cannot occur. Pathways sensitive to N17 H-*ras* are thus thought to depend on GEFs. Ras-dependent pathways that do not depend on GEFs may exist, however. For example, Ras may be activated by inhibiting GAP activity without the need for GEFs, a process which may therefore be insensitive to N17 H-*ras*.

Cdc25 and Scd25 from *S. cerevisiae* interact with human Ras[5], and must therefore bind to Ras proteins at highly conserved sequences. Mutational analysis indeed implicates conserved amino acids involved in the coordination of $Mg^{2+} \cdot GDP$, and the binding of the GDP purine ring[22], in addition to residues 73–75 (ref. 23) and 102–103 (ref. 24) as key elements in Ras activation.

A rasGEF has recently been shown to be responsible for activating Ras in response to epidermal growth factor (EGF). On binding EGF, the EGF receptor dimerizes and autophosphorylates tyrosine residues in its cytoplasmic domain, creating binding sites for Src-homology-2 (SH2) domains from other proteins. In humans, one such protein is the 'adapter protein' Grb2 (homologous to Sem-5 in *C. elegans* and Drk in *Drosophila*), which, in turn, binds to the rasGEF known as Sos1 via its Src. homology 3 (SH3) domains, which recognize proline-rich regions at the Sos1 C terminus (Fig. 2*a*)[25–27]. Through these protein–protein interactions, cytoplasmic Sos1 is brought to the plasma membrane, inducing Ras activation. Non-receptor tyrosine kinases may activate Ras in a similar way: these enzymes directly or indirectly phosphorylate the adapter protein Shc, which binds the Grb2–Sos1 complex in a manner similar to activated EGF receptor[28,19], presumably bringing the bound Sos1 to Ras at the plasma membrane. Whether activation of GTPases by GEF relocation is a general phenomenon remains to be seen. As Fig. 2 shows, only Sos proteins contain recognizable Grb2-binding domains, but other motifs (such as SH3 domains in Cdc25 and Scd25, or the pleckstrin homology regions in rasGRF, discussed below) could be involved in signal-dependent translocation.

Of the rasGEFs shown in Fig. 2, only Sdc25/Cdc25 and Sos/mSos form truly homologous sequence pairs, in the sense that their sequences can be meaningfully aligned along their entire lengths[5,30]. All nine proteins, however, share a domain of ~200 residues composed of three structurally conserved regions (SCRs) separated by more variable regions. A 13-residue consensus pattern contained within SCR3 is absolutely specific for known rasGEFs (see Fig. 2 legend). By analogy with the rasGAP 'catalytic' domain (see below), it is tempting to speculate that this region contains some of the critical determinants for rasGEF function. Four temperature-sensitive alleles of *cdc25* (*cdc25-1*, *-2*, *-5* and *-10*) indeed contain point mutations in SCR1 and SCR2 (ref. 31). A somewhat larger fragment of Cdc25 than that displayed in Fig. 2, however, is necessary for full biological activity[19].

ralGEF. In an increasingly common strategy, this GEF was cloned from a mouse cell line by the polymerase chain reaction using degenerate oligonucleotide primers derived from the conserved GEF domains of the yeast proteins Cdc25 and Ste6 (ref. 32). It is at least 20-fold more active on RalA and RalB than on members of the Ras, Rho/Rac and Rab GTPase families.

rapGEF. Cell polarity and budding in *S. cerevisiae* involves multiple GTPases of the Rap and Rho subgroups[33]. *BUD1* encodes a Rap-like protein, loss of which causes random budding. *BUD5* encodes a protein whose C terminus is homologous to the rasGEF domain common to Cdc25-like proteins (Fig. 2) and seems to be a GEF for Bud1 (ref. 5). Bud2 (a GTPase-activating protein, see below), together with Bud1 and Bud5, seem to comprise a functional module involved in bud site selection[34].

Genetic analysis implies that Bud5 can bind to *S. cerevisiae* RAS proteins, but it cannot replace Cdc25 function and thus seems to be an inefficient RAS activator[35]. As indicated by its ability to bind RAS proteins, Bud5's natural substrate, Bud1, is very similar to RAS in regions thought to be involved in GEF interactions (73–75, 102 and 103 of Ras, discussed above). We do not know which residues of Bud5 and Cdc25 make them specific for Bud1 and RAS.

A GEF specific for mammalian Rap proteins has not yet been identified. By analogy with BUD5 and ralGEF, rapGEF is likely to be a member of the Cdc25 family. Given Rap's ability to interfere with Ras signalling by blocking activation of Raf and the serine/threonine kinase cascade[36–39], it may be a protein of considerable importance.

rho/racGEFs. GEFs specific for Rac and Rho proteins are shown schematically in Fig. 3. The cell-cycle mutant CDC24 of *S. cerevisiae* seems from genetic analysis to have a defective GEF for the Rho-like protein CDC42 (ref. 33) (which is involved in bud assembly[33] and in determining cell shape, presumably through regulating the actin cytoskeleton), but this relationship has not been demonstrated biochemically. The related human oncoprotein Dbl (Fig. 3)[40], on the other hand, has been shown to act as a GEF for CDC42Hs[41] (the human homologue of CDC42, also known as G25K) and on Rho[127]. Furthermore, Dbl has recently been shown to bind to several Rac/Rho-like proteins (Rho, RhoC, Rac1, CDC42Mm) *in vitro*[42].

The capacity of Dbl to transform cells cannot yet be explained in terms of its GEF activity, because even permanently active mutants of CDC42Hs do not seem to be oncogenic (S. Munemitsu, unpublished). Perhaps CDC42Hs must cycle between its GDP- and GTP-bound states in order to function, but Dbl may also act as a GEF for other Rho-like proteins that can participate in transformation. Recently, the oncoprotein Ect2 has been shown to contain a Dbl-like GEF domain (Fig. 3)[42]. Ect2 expressed in insect cells showed no GEF activity, even though the recombinant protein (like Dbl) bound to Rac and Rho proteins, indicating that the 'Dbl domain' may not be a true GEF domain. Dbl's apparent GEF activity for CDC42Hs may thus be a fortuitous consequence of its interactions with other Rac and Rho proteins, or these GEF domains may require additional

factors to act on their substrates.

Efforts to show that purified Vav, Cdc24, Bcr and rasGRF are GEFs for Rac or Rho proteins have also been unsuccessful so far, and we are still unable to explain how rho/racGEF domains contribute to the biological activities of these proteins. Antibodies against Vav, however, have recently been used to show that Vav is associated with GEF activity in activated T cells. Surprisingly, this activity is directed against Ras, and Vav produced by *in vitro* translation generates rasGEF activity[128]. Vav must thus either associate with a GEF for Ras (a conclusion not fully supported by the data), or itself have rasGEF activity. The latter possibility would be unexpected, given the genetic and biochemical analysis of Vav-related proteins, but cannot be excluded.

The putative GEF domain of Cdc24-related proteins corresponds to a region of about 180 residues (Fig. 3) containing a 26-residue diagnostic motif. Like the rasGEFs, many of the rho/racGEFs contain additional interesting domains, including other GAPs and other GEFs, as well as SH2 and SH3 domains. RasGRF has recently been shown to contain two copies of a domain first described in pleckstrin[43], a protein of relative molecular mass 47K which is the major substrate for protein kinase C in platelets[44]. Pleckstrin homology (PH) domains have also recently been noted in Sos (Fig. 2), p120-GAP (Fig. 5) and other proteins (reviewed in ref. 45), but their function(s) are unknown. Vav contains a histidine- and cysteine-rich region closely related to the diacylglycerol (DAG)-binding domain of protein kinase C—a feature also shared by the rho/racGAP chimaerin (see below). This signature (Fig. 6) also occurs in the amino-terminal domain of Raf kinases which interacts directly with Ras[38,39].

ranGEF. The human gene *RCC1* encodes a 46K protein which was originally described as a regulator of chromosomal condensation[46]. Purification of RCC1 revealed an associated protein that was subsequently shown to be a GTPase of the TC4 family, now referred to as Ran (Ras-related nuclear protein). The RCC1/Ran complex can be dissociated with guanine nucleotides[47], indicating that it may be a stable, nucleotide-free intermediate formed during the process of GDP/GTP exchange (as discussed above for CDC25 and Ras)[24], and RCC1 indeed has subsequently been shown to promote nucleotide exchange

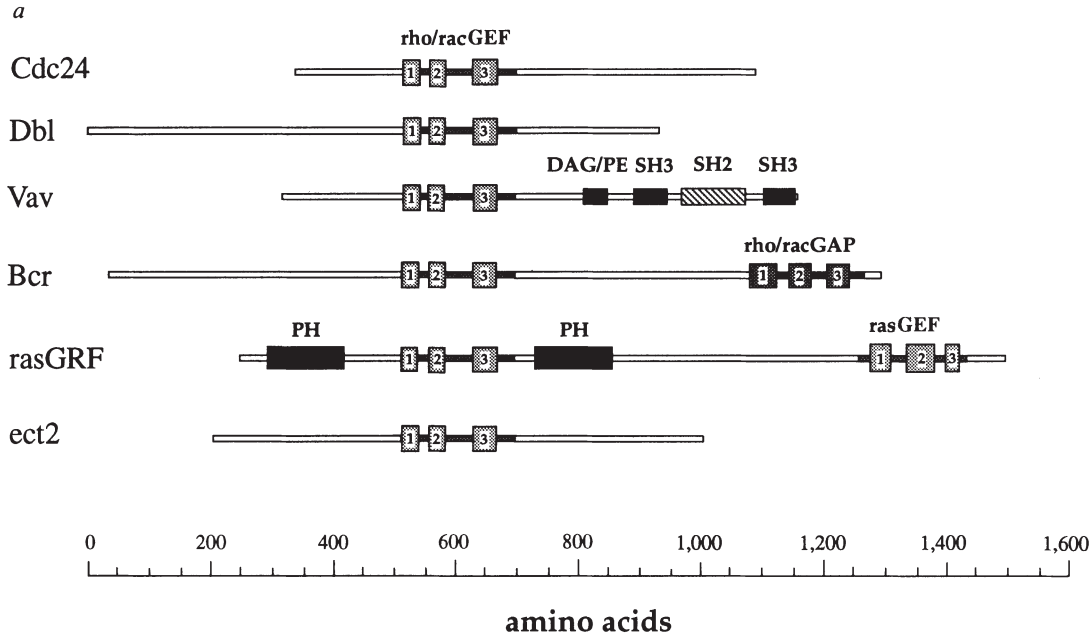
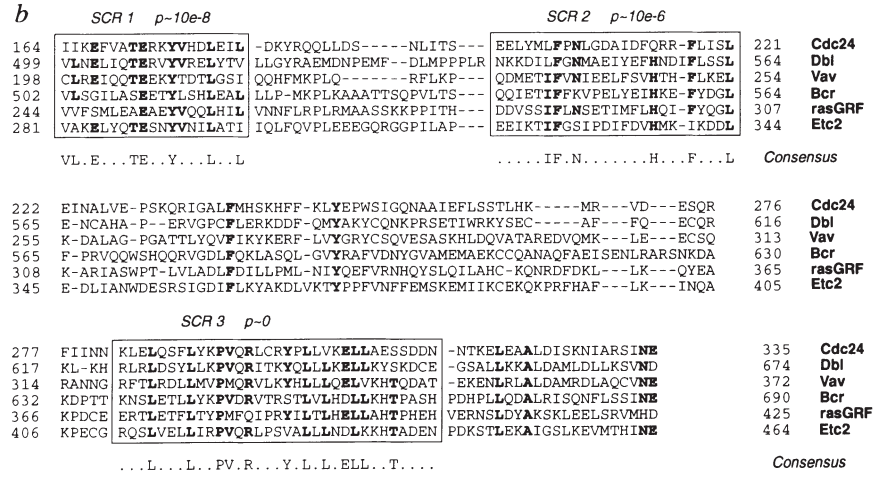


FIG. 3 Conserved domain in rhoGEF proteins. Programs and analyses were essentially the same as those in Fig. 2. Sequences of Cdc24 and protoVav are from *S. cerevisiae* and mouse, respectively; Dbl and Bcr are human sequences and rasGRF is from the rat (see Table 1). The regions that are most highly conserved among all five sequences represent the putative rhoGDS domain and a PROSITE signature (PDOC00605) for this domain is derived from SCR3: L-×(2)-[LIVMFYW]-×(2)-P-[LIVM]-×(2)-[LIVM]-×-[KR]-×(2)-L-×-[LIVM]-×-[LIVM](2)-(2)-[ST]. The DAG/PE and SH3-SH2-SH3 regions of Vav are according to ref. 126. The racGEF region in Bcr is described in detail in Fig. 6. The rasGNRP domain of rasGRF is described in detail in Fig. 2. For the alignment in *b*, significant similarity may extend another 80–100 residues beyond the last residues shown for Cdc24, Vav, Dbl and Bcr, but this additional region is not conserved in rasGRF. Cdc24, Dbl, Vav and Bcr are said to possess PH domains but, in comparison with those found in pleckstrin, p120GAP, rasGRF and Sos, the alignment scores are much less significant and not consistently distinguishable from background[45].



on Ran *in vitro*[47]. In *S. pombe*, RCC1 and TC4 homologues referred to as *pim1* and *spi1*, respectively, are involved in coupling the completion of DNA synthesis to initiation of mitosis[4].

RCC1 itself is composed almost entirely of seven repeats of about 60 residues each[46] and is unrelated in sequence to other known GEFs. A diagnostic signature for RCC1 repeats has been derived (PROSITE[48] PDOC00544), but the repeats seem not to be functionally interchangeable, as a single amino-acid substitution in the fourth repeat results in a mutant phenotype[49].

rabGEF. GEFs specific for members of the Rab family have been characterized biochemically[50] and molecularly[51]. The *S. cerevisiae* gene *DSS4-1* (dominant suppressor of sec4) encodes a 17K protein with GEF activity for sec4 (and to a lesser extent Ypt1) but not for rab3A. A rabGEF from rat brain was cloned by virtue of its ability to suppress the effects of a temperature-sensitive *SEC4* allele in yeast[52]. This protein (called Mss4 for mammalian suppressor of sec4) stimulates GDP release from Sec4, Ypt1 and rab3A. Neither Dss4 nor Mss4 has sequence similarities to other GEFs, but they do possess significant ($P \sim 0.09$), although very localized, similarities to each other[52]. The consensus sequence for the most highly conserved region is [LV]-K-[YF]-L-[IV]-C-A-D-C-x-G-P-I-G-x(2)-C.

smgGDS. Smg (small GTP-binding protein) GDS was originally described as a GEF for mammalian Rap proteins[53], but it also promotes nucleotide exchange on $p21^{K\text{-}ras}$ (although not in $p21^{H\text{-}ras}$), Rho and Rac proteins[54]. The broad specificity of smgGDS is surprising, as Ras and Rho proteins are only 35% identical in sequence, whereas the K-*ras*- and H-*ras*-encoded proteins are 85% identical. One clue to this unexpected range of specificity comes from the observation that smgGDS works efficiently only on isoprenylated proteins[22] and SmgGDS must therefore recognize features of the hypervariable carboxy terminus that include this modification. Ras and Rho/Rac proteins are modified by different isoprenoid moieties, however, as Ras sequences receive 15-carbon farnesyl groups whereas Rho/Rac proteins receive 20-carbon geranylgeranyl groups (see below), and it is difficult to explain how SmgGDS distinguishes farnesylated K-*ras* protein and geranylgeranylated Rho from farnesylated H-*ras* protein.

These properties suggest that SmgGDS is a distinct type of Ras regulator, having little in common with other classes of GEFs, all of which react with non-prenylated forms of their targets and exhibit subfamily specificity. SmgGDS has no sequence similarity to any other GEF.

Guanine-nucleotide-dissociation inhibitors (GDIs)

rabGDI. A protein first isolated from bovine brain[5] affects the rate of GDP dissociation from Rab3A (also known Smg p25A). This protein, here referred to as rabGDI, forms a 1:1 complex with Rab3A, inhibiting GDP/GTP exchange and preventing the GDP-bound form from binding to membranes (Fig. 4)[5]. Both of these activities seem to depend on the C-terminal geranylgeranylation of Rab3A (ref. 5) and rabGDI may only recognize the prenylated form of this GTPase. Its tissue distribution indicates that Rab3A may be involved in the regulation of secretion at brain synapses and from endocrine and exocrine cells[55]. A closely related GDI has also been identified in rat liver cytosol, however, and this GDI may be a tissue-specific regulator of the 24K liver GTP-binding protein G[5].

The *Drosophila* homologue of rabGDI is one of three abundant proteins whose isoelectric points are shifted in the developmental and mitotic mutant *quartet* (ref. 56). Several aspects of the quartet phenotype indicate that this GDI, like the mammalian rabGDI, may play a role in secretion or endocytosis.

A gene encoding a GDI has been identified on chromosome 5 of *S. cerevisiae* (J. Mulligan, personal communication). A second, distinct yeast protein, MRS6, shares highly conserved sequence motifs with both GDI proteins and the human choroideraemia gene product, defects in which are responsible for an X-linked form of hereditary blindness[57]. The choroideraemia gene product, known as $p95^{CHM}$, has not been shown to possess GDI activity, but is a subunit of a brain enzyme, geranylgeranyl transferase[58] (Table 2), which adds prenyl groups to the carboxy termini of Rab proteins (Fig. 4; see below). The conserved motifs may thus be responsible for rab3A recognition and binding rather than GDI activity *per se*[57].

rhoGDI. RhoGDI was first identified as a factor capable of inhibiting dissociation of GDP from post-translationally modified Rho proteins[5,59]. Its biological role was suggested by its ability to remove Rho proteins from cellular membranes in cell-free systems[5], indicating that it might regulate the proportion of the available Rho proteins associated with the membranes, or facilitate movement of Rho proteins from one membrane compartment to another. Without such a factor, Rho proteins are expected to remain firmly attached to the membranes[59], as the geranylgeranyl group at their C-termini provides a much more secure attachment to membranes than the farnesyl groups of Ras proteins, for example.

Rac proteins bound to rhoGDI have also been identified as components of the NADPH oxidase system which generates oxygen radicals in activated phagocytes (reviewed in ref. 10). Rac and rhoGDI form a heterodimer called σ1 which is required for oxidase stimulation *in vitro*, along with two other cytosolic factors ($p67^{phox}$ and $p47^{phox}$), both of which contain SH3 domains[61]. These three components assemble into a membrane-bound complex which, together with a cytochrome *b*, uses electrons from NADPH to generate superoxide anions[10]. Recombinant Rac proteins, in their GTP-bound state, can replace the requirement for Rac+rhoGDI in this system, indicating that rhoGDI may recognize the GTP-bound form of Rac and protect it from racGAPs. This has now been shown for both Rac and other GTPases[48,62-64], raising the possibility that rhoGDIs play a physiological role as GAP inhibitors. If so, it will be of interest

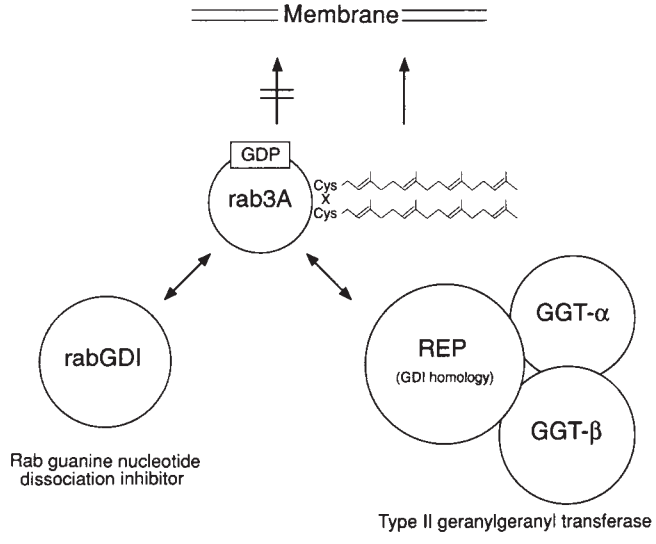


FIG. 4 Interactions between Rab3A, rabGDI and geranylgeranyl transferase. Rab3A (also known as smg p25A) is a Ras-like GTPase with an approximate M_r of 25Kd. RabGDI is an ~50K protein[5] that binds to the prenylated form of Rab3A·GDP and inhibits both the GDP/GTP exchange reaction (see Fig. 1) and the interaction of Rab3A with membranes[5]. The carboxy terminus of Rab3A is modified by the addition of two 20-carbon geranylgeranyl groups[5] in a reaction catalysed by the heterotrimeric enzyme geranylgeranyl transferase (Rab GGTase or GGTase, type II)[109,111]. Rab GGTase from rat brain is an ~330K complex of three subunits that is dissociable into two components (Table 2). Component A (or 'Rab escort protein', REP) is a homologue of the human chorioderaemia gene product, $p95^{CHM}$ (see text). Component B (GGT-α+GGT-β) is the catalytic component.

TABLE 1 Some proteins that regulate Ras-like GTPases

Gene or protein name	Organism(s)	Size in amino acids	Database accession no.	Comments	Specificity +	Specificity −
(a) Guanine-nucleotide-exchange factors						
RasGEF						
Ras-GRF	Rat, mouse Human	1,244 Partial	SP:P28818 GB:M91815	Cloned by sequence homology to Cdc25 (refs 117, 118), and by genetic complementation GEF activity *in vitro*. Expressed in brain.	H-*ras*, N-*ras* (K-*ras* not tested)	ralA, CDC42HS
Sos	Human Mouse *Drosophila*	1,333 1,336 1,595	GP:L13857 SP:P27671 SP:P26675	Cloned by sequence homology to Cdc25 (ref. 30). mSos-1 67% identical to mSos-2. Widely expressed during development and in adult tissues. GEF activity *in vitro*. Binds Grb2/Sem5. Connects Ras to tyrosine kinases. *Drosophila* Son of Sevenless: necessary for signalling from Sevenless receptor to Ras1 (ref. 120).	H-*ras*	R-*ras*
Ste6	*S. pombe*	911	SP:P26674	Essential for Ras1 function in mating pathway. Presumed GEF.	?	?
Cdc25	*S. cerevisiae* *S. kluyveri* *C. albicans*	1,589 1,333	SP:P04821 GP:M82964 GP:M94160	First rasGEF identified[4]. Essential for RAS activation, viability. GEF activity *in vitro*.	RAS2 (others not tested *in vitro*	?
Sdc25	*S. cerevisiae*	1,250	SP:P14771	Non-essential. Cannot replace Cdc25 (ref. 4). GEF activity *in vitro*.	RAS1, RAS2, H-*ras*	Rap1A, Rap2, RalA, YPT1, Ryh1
RalGEF						
	Mouse, rat	852	GP:L07924		RalA, RalB	Ras, Rap, Rho, Rab
RapGEF						
Bud5	*S. cerevisiae*	538	SP:P25300	GEF function presumed from genetic analysis of budding phenotype.	BUD1? (inferred from genetics)	RAS1, RAS2? (inferred from genetics)
Unclassified						
Lte1	*S. cerevisiae*	1,186	GB:L20125	Loss of function results in cold sensitivity[4]. Presumed GEF.	?	?
RhoGEF						
Cdc24	*S. cerevisiae*	736	SP:P11433	GEF activity not yet shown, inferred from genetics.	Cdc42? from genetics	Ras1, Ras2? (from genetics)
Dbl	Human	925	SP:P10911	Truncated version transforms cells. GEF activity shown *in vitro*, but with pure components. No requirement for lipid modification of substrate.	Cdc42HS, Rho (poor substrate)	H-*ras* Rap1A Rac
Vav	Human, etc.	845	SP:P15498	Truncated version transforms cells. Associated with GEF activity for Ras, not Rac/Rho. Phosphorylated on tyrosine residues in activated cells.	Ras?	?
Bcr	Human	1,271	SP:P11274	Putative kinase. GEF activity not yet shown. Forms chimaera with abl in CML, ALL. GEF domain retained in CML chimaera (p210 protein), lost in ALL (p185).	?	?
Ras-GRF	Rat, mouse	1,244	SP:P28818	GEF activity not yet shown.	?	?
Ect2	Mouse	738	GP:L11316	Oncogene product.	?	?
RabGEF						
Dss4 MSS4	*S. cerevisiae* Rat	143 123	GB:X70495 GB:S54294	Protein sequences 16.7% identical; *DSS4-1* allele has stop codon replacing Asp 108.	Sec4 > Ypt1 Sec4 > Ypt1, Rab3A	Ras2, rab3A Ras2
RanGEF						
RCC1 RCC1 RCC1 BJ1 Srml/Prp20 Piml	Human Hamster *Xenopus* *Drosophila*, *S. cerevisiae*, *S. pombe*	421 421 424 547 482 539	SP:P18754 SP:P23800 SP:P25183 SP:P25171 SP:P21827 SP:P28745	GEF activity *in vitro*. *Xenopus*, *Drosophila*, *S. cerevisiae* genes complement hamster ts mutant.	Ran	Ras
Multispecific						
smgGDS	Human, etc.	558	GPX63465	Requires lipid modification of substrate.	RhoA, Rac1 K-*ras*	H-*ras*, Rab3A, Rab11, Sec4
(b) Guanine-nucleotide-dissociation inhibitors (GDIs)						
Rab/Ypt1						
RabGDI	Human, etc.	447	SP:P21856	Requires lipid modification. Regulates association of Rab with membranes.	Rab3A, Rab6, Rab11, Rab9, Sec4	H-*ras*, K-*ras*, Rap1A, RhoB
	Drosophila *C. elegans* *S. cerevisiae*	448 444 452	GP:L03209 GP:U00002 —	GDI activity inferred by homology. GDI activity inferred by homology. GDI activity inferred by homology.		

TABLE 1—continued

Gene or protein name	Organism(s)	Size in amino acids	Database accession no.	Comments	Specificity +	Specificity −
Mrs6	*S. cerevisiae*	603	GP:M90844	Shares conserved motifs with above sequences; not globally homlogous.	?	?
p95-CHM	Human, etc.	656	SP:P26374	Choroideraemia gene; Rab 'escort' subunit of type II geranylgeranyl transferase.		
Rho/Rac						
RhoGDI	Human, etc.[94,95]	294	SP:P19803	Requries lipid modification.	Rho, Rac	Ras, Rab
	C. elegans	(partial)	GB:Z14649		?	?
(c) GTPase-activating proteins						
RasGAP						
Neurofibromin (NF1)	Human, etc. *Drospophila*	2818 >2,600	SP:P21359 —	Defective allele responsible for neurofibromotosis disease. Tumour suppressor function in NF1-related malignancies. Dominant mutants in some solid tumours. Major Ras regulator in neural cells (myeloid cells?). Strongly inhibited by lipids, arachidonic acid[5]. Expressed ubiquitously.	H-*ras*, N-*ras* (K-*ras*, RAS1, RAS2 (*S. cerevisiae*) binds oncogenic Ras proteins without stimulating GTPase. Rap1A not tested.	Rho, Rac, Rab
p120GAP	Human, etc.	1047	SP:P20936	Binds tyrosine phosphoproteins (PDGF-R EGF-R, p62 (ref. 121) via SH2 domains. Binds p190 (rho/racGAP). Splicing variant p100GAP expressed in placenta. Inhibited in activated T cells. Binds to Ras at 'effector' domain. Binds Rap1, but does not stimulate Rap1 GTPase activity[5]. Necessary for oocyte maturation[77].	H-*ras*, N-*ras* K-*ras*, R-*ras*, RAS1, RAS2 (*S. cerevisiae*) ras1 (*S. pombe*). Binds oncogenic Ras proteins, Rap1A, without stimulating GTPase.	Rho, Rac, Rab
Gap1	*Drosophila*	1165	GP:M86655	Negative regulator of Ras1 in sevenless pathway: no effector function in this pathway[81]. No SH2/3 domains.	Ras1	—
Ira1	*S. cerevisiae*	2938	SP:P18963	Negative regulator of RAS1 and RAS2. No effector function in cyclase pathway[4].	RAS, RAS2	H-*ras*? Rho, Rac, Rab
Ira2	*S. cerevisiae*	3079	SP:P19158	Negative regulator of RAS1 and RAS2. No effector function in cyclase pathway[4].	RAS, RAS2	H-*ras*? Rho, Rac, Rab
Sar1/gap1	*S. pombe*	765	PI:A40258	Negative regulator or ras1 in mating pathway. No effector function in this pathway.	Ras1 (*S. pombe*); RAS1, RAS2 (*S. cerevisiae*)	?
Bud2	*S. cerevisiae*	1,104	GB:L19162	Bud2 mutant shows random budding but no growth defect. Substrate Bud1, is most similar to Rap (ref. 34).	Bud1	
RapGAP						
Rap1GAP	Human	663	PI:A39897	No detectable homology with other GAPs; substrate for PKA and p34 (ref. 84).	Rap1A, Rap2	Ras, Rho, Rac
Rho/racGAP						
Bcr	Human	1,271	SP:P11274	GAP activity *in vitro*. GAP domain removed in bcr/abl chimaeras. Putative rhoGNRP domain.	Rac, CDC42Hs	Rho, Ras
n-Chimaerin	Human, etc. Canary	299 299	SP:P15882 GP:S98891	Expressed in brain and testis. Contains DAG/PE signature. GAP activity *in vitro*.	Rac	Rho, CDC42Hs, Ras
rotund locus	*Drosophila*	384	GB:M99702	Cysteine-rich domain.		
p190	Rat	1,513	GP:M94721	Associates with p120GAP. Substrate for tyrosine phosphorylation. GRF-1 domain and purine nucleotide binding motif.	Rac, Rho, CDC42Hs	Ras
GRB-1/p85a	Human, etc.	724	SP:P27986	Associates with receptor tyrosine kinases via SH2 domains, and with catalytic subunit of PI3 kinase. No GAP activity *in vitro* (yet).		
3BP-1	Mouse	(partial)	EN:253168	SH3 binding protein. GAP activity not shown yet.		

Listed are various proteins discussed in the text along with the species in which they have been identified, their size in amino-acid residues and the database accession numbers for those sequences used in analyses presented in Figs 2–6. Where the 'human, etc.' designation is used, homologous sequences have been isolated from other mammals, usually rodents or cows. The database designations are as follows: SP: = SWISS-PROT (release 26.0); GB/GP: = GenBank/GenPept release 77.0); PI: = PIR International (release 37.0); and EN: = NCBI Entrez database. When a sequence was present in more than one database, the SWISS-PROT entry was preferred because of its extensive annotation of protein sequence features (for example SH2/3 domains and the presence of internal repeats) and its comprehensive cross-references to other databases including the PROSITE database of protein sequence motifs[22].

to determine how GAPs gain access to rac·GTP, and whether interaction with $p67^{phox}$, $p47^{phox}$ or cytochrome *b* displaces rhoGDI.

GTPase-activating proteins (GAPs)

rasGAP. Ras proteins have a very low intrinsic GTPase activity and their inactivation is dependent on GAPs *in vivo*. Oncogenic mutants of Ras proteins, for example, at positions 12, 13 or 61, are resistant to GAP-mediated GTPase stimulation and are constitutively locked in their active, GTP-bound states[65,66].

The first GAP characterized at the molecular level was p120-GAP (ref. 5). In addition to a hydrophobic amino terminus, two SH2 domains[61] and one SH3 domain[67] (Fig. 5), it contains a PH domain and a region similar to the CalB region of phospholipase A_2 (ref. 68). The last indicates that it may undergo calcium-dependent translocation to cellular membranes under some circumstances.

A ~250-residue domain towards the C terminus of the protein is responsible for GTPase stimulation (Fig. 5)[4]; detailed alignment studies of this domain have been published elsewhere[69]. This domain, when isolated from its parent sequence, stimulates Ras GTPase activity, but with kinetics different from those of full-length p120-GAP[70]. Significant contacts may therefore occur between distal regions and this 'catalytic domain' itself, although their significance is unknown.

Two models have been advanced to describe the consequences of these interactions. In the first model, p120-GAP undergoes a conformational change on binding Ras that affects its interactions with other proteins[71]. Ras would thus control p120-GAP's function, and the duration of its effect would be determined by the rate at which it is inactivated by p120-GAP, and oncogenic Ras mutants resistant to GAP-mediated GTP hydrolysis would activate GAP persistently. In this model, GAP p120-GAP serves as a Ras effector. p120-GAP SH2/3 domains separated from the C-terminal Ras-binding domain are indeed able to affect transcription, and there is evidence that p120-GAP SH3 domains are essential components of Ras downstream signalling pathways[72].

In the second model, interactions between p120-GAP and other proteins affect p120-GAP activity, thereby determining the level of active Ras. Evidence for and against these models has been discussed elsewhere[66,73], but it is by no means clear that GAPs are indeed biologically relevant Ras effectors. In support of the effector model, phospholipase C_β and the γ-subunit of phosphodiesterase (effectors for the heterotrimeric G proteins G_q and transducin, respectively) are indeed known to be GAPs[74,75]. Recent evidence that Ras and the serine/threonine kinase Raf interact directly[38,39], however, makes the latter a much stronger candidate for a meaningful Ras effector, as it is known to be activated by Ras and, when active, can replace Ras in some contexts[72]. It will be of considerable interest to determine which of the many reported effects of Ras are mediated by Raf, and whether alternative effector pathways exist.

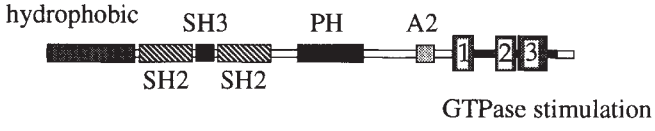


FIG. 5 Domains in p120-GAP. In contrast to other rasGAPs (Table 1c), nearly all of the 1,044 residues of this prototypic GTPase-activating protein have known functions, or suspected functions based on sequence similarities to other proteins. The hydrophobic and SH3-SH2-SH3 regions are as reviewed in ref. 5. The small rectangle labelled 'A2' is a putative CaLB motif believed to be a phospholipid-binding sequence involved in the Ca^{2+}-dependent translocation of certain regulatory proteins to membranes[68]. A pleckstrin homology (PH) domain has recently been described[45]. Detailed alignment studies of the GTPase stimulation domain have been published elsewhere[69,82]. Briefly, this rasGAP 'catalytic' domain consists of three highly conserved sequence blocks, the third of which contains a diagnostic PROSITE signature: F-L-R- × (3)-P-A- × (3)-P (PDOC00438).

S. cerevisiae contains two rasGAP proteins, IRA1 and IRA2 (ref. 5), which contain domains homologous to the C terminus of p120-GAP[4,5]. In wild-type cells, RAS proteins are generally inactive[4], but in the absence of either *IRA* gene product, they accumulate in their GTP-bound state, becoming hyperactive and leading to overproduction of cyclic AMP[73]. In yeast, at least, rasGAPs are therefore not Ras effectors, serving only as negative regulators.

Neurofibromin (NF1), the human protein defective in von Recklinghausen neurofibromatosis, contains a domain homologous to the 'catalytic' domains of p120-GAP, IRA1 and IRA2 (ref. 5). Like p120-GAP itself, neurofibromin possesses GTPase-activating activity *in vitro* and can complement the loss of IRA function in *S. cerevisiae*[5], indicating that it may be the mammalian homologue of IRA1 and IRA2. These proteins share regions of significant similarity outside the GAP-related domain[74,75], and presumably perform similar functions. p120-GAP may have evolved later along with the tyrosine kinase signalling pathways in which it is thought to play a key role.

The roles of neurofibromin and Ras in the pathology of neurofibromatosis are under intense investigation. Patients with von Recklinghausen neurofibromatosis due either to a defective *NF1* allele inherited from an affected parent, or to sporadic mutations in the *NF1* gene, display abnormalities of neural crest cell growth. Two (non-exclusive) models have been discussed to account for their phenotype. In the first, loss of one functional neurofibromin allele in neural crest cells, where it is the major Ras regulator, leads to the accumulation of GTP-bound Ras, just as loss of one *IRA* gene allows active RAS to accumulate in *S. cerevisiae*. The active Ras may contribute directly to cell proliferation, as shown in malignant Schwannomas from neurofibromatosis patients[76,77]. The second possibility is that neurofibromin plays a direct role in cell differentiation or growth control, so that loss of one allele allows defective differentiation and uncontrolled growth. Active Ras levels in neuroblastomas lacking one or both neurofibromin alleles are indeed unaltered, indicating that selection for loss of neurofibromin in these tumours was unrelated to its effects on Ras[78].

The role of neurofibromin may be more complex than either of these models suggests. Mutant *NF1* alleles are associated with sporadic cancers unrelated to neurofibromatosis or neural crest tissues[79]. These mutations do not seem to be loss-of-function mutations, as the same codon is altered in each of three cases, and it is difficult to explain how one defective neurofibromin allele could contribute to development of non-neural crest tumours in sporadic cases, but not in neurofibromatosis patients. Moreover, oncogenic Ras mutations are rare in sporadic tumours derived from the neural crest, indicating that raising active Ras levels by loss of neurofibromin function is somehow different from raising levels by Ras mutation. By contrast, juvenile myeloid leukaemias are associated with loss of both normal *NF1* alleles in neurofibromatosis patients, but with *ras* mutations in sporadic cases; this is one of the few cases in which neurofibromin loss is biologically equivalent to oncogenic Ras activation[80].

Drosophila contains a protein, 70% identical to neurofibromin (Table 1 and A. Bernards, personal communication), in addition to a distinct rasGAP (referred to as GAP1) which is a component of the Sos tyrosine kinase/Ras1 signalling pathway[81]. Loss of GAP1 stimulates Ras1 function, indicating that it is not an effector of Ras1 action. GAP1 has no SH2 or SH3 domains, and does not seem to be related to neurofibromin outside the rasGAP domain.

Unlike the *RAS* genes of *S. cerevisiae*, the *ras1* gene of the fission yeast *S. pombe* does not act through adenylyl cyclase and thus more closely resembles *ras* genes in mammals. To date, a

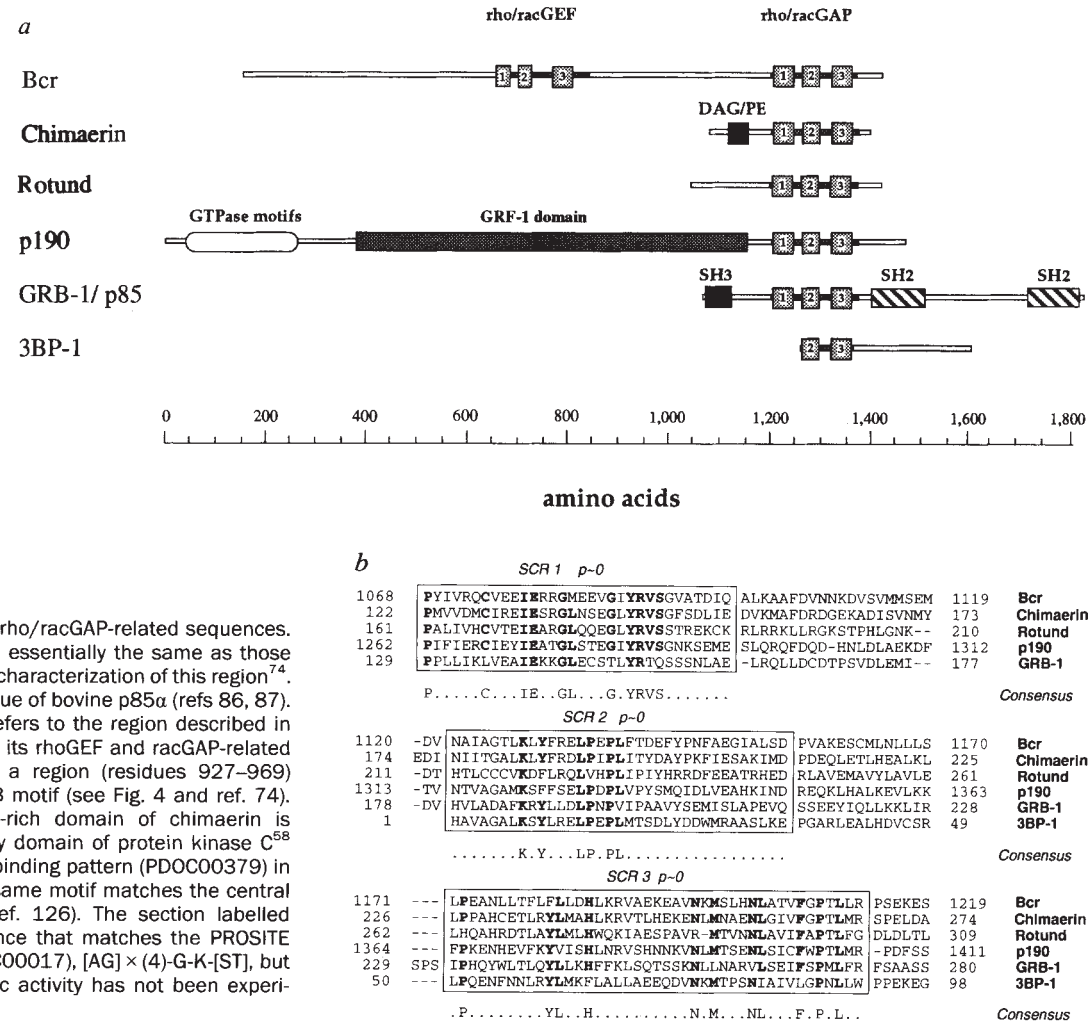


FIG. 6 Conserved domain in rho/racGAP-related sequences. Programs and analyses were essentially the same as those in Fig. 2 and follow an earlier characterization of this region[74]. GRB-1 is the human homologue of bovine p85α (refs 86, 87). The rhoGEF domain of Bcr refers to the region described in detail in Fig. 3. In addition to its rhoGEF and racGAP-related domains, Bcr also contains a region (residues 927–969) strongly resembling the CaLB motif (see Fig. 4 and ref. 74). The histidine- and cysteine-rich domain of chimaerin is homologous to the regulatory domain of protein kinase C[58] and is listed as the DAG/PE-binding pattern (PDOC00379) in the PROSITE database; this same motif matches the central region of Vav (Fig. 3 and ref. 126). The section labelled 'GTPase?' contains a sequence that matches the PROSITE ATP/GTP-binding motif (PDOC00017), [AG] × (4)-G-K-[ST], but ATP/GTP binding or hydrolytic activity has not been experimentally demonstrated.

single rasGAP gene called *sar1* or *gap1* has been identified in *S. pombe*. Although it is not globally homologous to any other known rasGAP, it does contain a related 'catalytic' domain, and can complement loss of IRA1 or IRA2 function in *S. cerevisiae*[82]. It is not a *ras1* effector.

Another protein from *S. cerevisiae*, RPI1, may inhibit Cdc25 (ref. 83). Overexpression of this protein inhibits normal Ras function, but not that of activated Ras. This effect depends on functional *IRA* genes, however, indicating that it may be mediated by these negative regulators. Unfortunately, the predicted amino-acid sequence of RPI1 offers no clues as to its function.

rapGAP. Rap1A is 53% identical to Ras and, like Ras, binds to p120-GAP[28] and to raf-1 (refs 36–38) by its effector-binding domain. Although Rap1A binds p120-GAP, its GTPase activity is not enhanced by this interaction, and another protein, rap1GAP, is responsible for Rap1A GTPase activation[5]. rap1GAP is unrelated to rasGAP or any other known protein, although it contains several sites for phosphorylation by Cdc2 and cyclic AMP-dependent kinases[84]. This total lack of similarity with rasGAP is surprising in view of the close relationship between Rap and Ras proteins. Ras proteins, like most GTPases, depend on a glutamine residue at position 61 (or equivalent) for intrinsic or GAP-mediated GTP hydrolysis, however, whereas Rap1 has a threonine at this position, and the GAP activities of rap1GAP and p120-GAP may therefore have different mechanisms.

rho/racGAPs. A mammalian GAP specific for Rho has been purified and shown to contain a region related to the C-terminal domain of Bcr (ref. 5), a putative rhoGEF mentioned above, and to a protein from human brain called n-chimaerin[85]. Bcr and n-chimaerin indeed stimulate GTP hydrolysis by the Rho-like proteins Rac1 and Rac2, but not by Rho proteins themselves; as predicted, this activity is mediated by the C-terminal 401 amino acids of Bcr[5]. This domain bears no significant resemblance to rasGAP or rap1GAP. A similar domain has also been noted in two 85K proteins (also known as GRB-1) associated with phosphatidylinositol 3-kinase (PI3-kinase) (Fig. 6)[74,86,87], although these proteins have not been shown to have GAP activity. In addition to its C-terminal rhoGAP domain, chimaerin also contains an N-terminal DAG-binding motif. Canary chimaerin messenger RNA encodes a protein 96% identical to its human counterpart; it is enriched in brain areas associated with complex learned behaviour, and is variably expressed in the song control circuit[88].

A multi-domain protein termed p190, which binds to p120-GAP and regulates its activity[60], contains a central domain related a putative DNA-binding transcriptional repressor[89]. A 145-residue region at its C terminus is related to rhoGAPs and the protein is indeed a potent GAP, for all members of the Rho/Rac family with which it has been tested (Table 1 and ref. 115).

Finally, the *rotund* locus of *Drosophila* encodes a protein with a putative rho/racGAP activity[90]. Mutations in this locus result

in shortened appendages through localized cell death in imaginal disks.

ralGAP. A GAP activity specific for Ral proteins (50% identical to the H-*ras*-encoded protein) has been reported[5]. The protein responsible has not been sequenced, but its properties differ from those of other known GAPs, as it sediments on density gradients, for example, with an apparent M_r of more than one million. It does not interact with Ras.

rabGAP. GAP activities specific for individual members of the Rab family (Ypt1, ref. 91 and Sec4, ref. 92) are known, and a complementary DNA for a Rab-specific GAP has been cloned[93]. This protein, GYP6, shows no similarity to other GAPs and is remarkably specific for Ypt/rab6. Individual GAPs for each member of the Rab family may therefore exist, perhaps playing a role in membrane recognition by these proteins.

Prenyltransferases

Ras-like GTPases are targeted to the membranes where they act by the posttranslational attachment of isoprenoid lipids (or prenyl groups)[94,95]. Prenylation involves covalent thioether linkages of farnesyl (15-carbon) or geranylgeranyl (20-carbon) groups (derived from the mevalonate → cholesterol pathway) to cysteine residues near the proteins' C termini (see, for example, Fig. 4). These reactions are catalysed by three prenyltransferases (Table 2) which differ in their isoprenoid substrates and protein targets. Farnesyltransferase (FTase) modifies Ras family members with Carboxy-terminal tetrapeptides of the form Cys-A-A-X, where 'A' represents an aliphatic amino acid and 'X' is usually a methionine or serine residue. Type I geranylgeranyltransferase (GGTase-I) also recognizes a CAAX motif, but prefers a leucine residue in the 'X' position; its substrates include members of the Rho/Rac and Rap families. Type II GGTase (GGTase-II, also referred to as RabGGTase), on the other hand, recognizes carboxy-terminal motifs of the form Cys-X-Cys or Cys-Cys, typical of Rab proteins.

The α-subunits of prenyltransferases (Table 2) have all evolved from a common ancestral gene and contain a repetitive sequence motif with an invariant tryptophan residue and other conserved amino acids[74,96–98] that are important for function[99]. Prenyltransferase β-subunits are also ancestrally related and contain internal repeats[74,96–98]. In this case, however, the repetitive motif contains a Cys-His-Cys triad that may account for Zn^{2+} binding[97,100].

Biochemical similarities and differences among the prenyltransferases are described below.

CAAX prenyltransferases. The prenyltransferases have been most thoroughly studied so far in rat brain and *S. cerevisiae* (Table 2). Rat FTase is a heterodimer consisting of α- (~49K) and β- (~46K) subunits[72]. Chemical cross-linking studies[101] have shown that Ras binds FT-β in a reaction requiring Zn^{2+} (ref. 100). Prenyl pyrophosphates can bind (probably to the α-subunit) without divalent actions, but Mg^{2+} is necessary to transfer the prenyl group to the peptide acceptor[100]. The *S. cerevisiae* genes *RAM2* and *RAM1* encode the yeast homologue of FT-α and FT-β, respectively[102], and the yeast enzyme is functionally similar to the mammalian enzyme[102]. FTase has recently attracted much attention because CAAX-mimetic drugs inhibiting this enzyme show promise as anticancer agents[104,105].

Interestingly, the FT-α subunit in both yeast[106] and mammals[107] combines with a different β-subunit to form GGTase-I (Table 2). The β-subunit in *S. cerevisiae* is the product of the *CDC43* (or *CAL1*) gene[50] involved in cell-cycle progression, cell polarity and budding[108].

Rab geranylgeranyltransferase. GGTase-II consists of three subunits[58,109]. Unlike FTase, it is inhibited by Zn^{2+} and recognizes other domains of the target protein in addition to its carboxy-terminal substrate[58,109].

Native GGTase-II is a 330K complex (Fig. 4) which can be dissociated into two components[58,109]. Component B (the 'catalytic component'[110]) is a heterodimer of 38K and 60K subunits analogous to the other two prenyltransferases (Table 2). These mammalian α- and β-subunits are clearly homologous to the proteins MAD2 and BET2 from *S. cerevisiae*[111], respectively. BET2 is required for the membrane attachment of the Rab subfamily GTPases, Ypt1 and Sec4 (ref. 112), whereas MAD2 is involved in mitotic feedback control[74,97]; both have recently been shown to be subunits of the yeast GGTase-II[113,114].

Component A of the rat brain GGTase-II is the rat homologue of the human choroideraemia gene product (see section on rabGDI above). It may function as a 'Rab escort protein', removing prenylated Rab3A from the catalytic subunits so that further proteins can be modified[110]. Preliminary evidence suggests that the yeast GGTase-II also requires an A component for maximal activity[113] and its sequence indicates that Mrs6 (ref. 57) is a likely candidate.

TABLE 2 Classes of known protein prenyltransferases

Prenyltransferase		Quaternary structure α-subunit	β-subunit	
CAAX farnesyl-transferase	Yeast	RAM2 (316 residues)	RAM1 (431 residues)	
	Mammals (rat brain)	FT/GGT-α (377 residues)	FT-β (437 residues)	
CAAX geranylgeranyl-transferase (GGTase-I)	Yeast	RAM2 (316 residues)	CDC43 (376 residues)	
	Mammals (rat brain)	FT/GGT-α (377 residues)	GGT-β (377 residues)	
Rab geranylgeranyl-transferase (GGTAse-II)	Yeast	MAD2 (290 residues)	BET2 (322 residues)	MRS6? (603 residues)
	Mammals (rat brain)	α (567 residues)	β (331 residues)	Choroideraemia gene product (650 residues)
		Component B catalytic component		Component A 'Rab escort protein'

The prenyltransferase subunits from *S. cerevisiae* and rat brain are the most thoroughly studied thus far and are used here for comparitive purposes. However, a cDNA clone of FT-α from bovine brain has also been described[123] and α-subunit homologues have been identified in cDNA libraries of human brain (hippocampus) and the nematode worm *Caenorhabditis elegans*[96,97] using the EST database[124]. A *C. elegans* Bet2 homologue has also been discovered in this manner[96,97]. Database accession numbers, along with detailed analyses of these sequences are given in references[74,96,97,111], except for rat GGTase-Iβ (ref. 107) whose GenBank accession number is L24116.

Specificity and crossregulation of Ras regulators

The different families of proteins within the Ras superfamily are thought to be involved in distinct physiological processes, as discussed above, and the individual members of each family presumably perform specialized aspects of these functions. The specificity with which regulators interact with individual members of a family may reflect the degree to which the member's functions are specialized. Amongst the Ras family, both effectors and regulators show remarkably little specialization. Oncogenic mutants of H-*ras*, K-*ras* and N-*ras* have comparable effects, and their regulators show little preference for one protein rather than another. p120-GAP, for instance, acts on all known Ras family members, including those of yeast, and on the more distantly related mammalian R-*ras*[5]. Neurofibromin is slightly more selective, with different binding constants for H-*ras* and N-*ras* p21 (ref. 5), but can function effectively on the RAS proteins of *S. cerevisiae*, while the mammalian rasGEF, RasGRF, acts on both H-*ras* and N-*ras* encoded proteins, and related mammalian rasGEFs have been cloned on the basis of their ability to activate yeast RAS proteins. All these proteins interact with the highly conserved GTPase domain of their targets, perhaps accounting for their inability to distinguish between Ras family members. Presumably, the 'variable regions' that distinguish each Ras protein are involved in more subtle aspects of their function which have so far escaped detection.

Amongst the regulators of Rho family members, p190 is a promiscuous GAP for all Rho-like proteins[115] whereas Bcr is Rac-specific, and the related protein, GRB-1 (the p85a subunit of PI3'kinase), fails to act on any known Rho/Rac-like protein. This variation in specificity indicates that these proteins interact with 'variable' regions of their substrates instead of the more highly conserved GTPase domain. While the specificity of these GAPs is consonant with their involvement in distinct biological pathways (Rac and ruffling; Rho and stress fibres), the promiscuity of p190 is harder to explain. Presumably some signals demand a coordinated response from all Rho/Rac proteins, whereas others are selective. The resolution of this issue will involve some fascinating biology.

Whereas these variations in regulator specificity can be reconciled with the specificity of their targets' effector functions, the existence of regulators which act on GTPase of different groups is harder to explain. Such proteins may serve to interconnect different regulatory pathways. For example, the mammalian rasGEF, rasGRF, also contains a GEF domain for an unidentified Rho-like protein (Figs 2, 3), and Bcr contains a potential GEF domain for Rho proteins in addition to its racGAP activity. Similarly, smgGDS activates specific members of the Rap, Ras and Rho subgroup.

The possible complexity of these pathways is increased further by interactions between different regulatory proteins themselves. p120-GAP binds to the rhoGAP p190, for instance, linking Ras mto Rac and Rho proteins, and its N-terminal domain has effects on the cytoskeleton that may be mediated by p190 (ref. 116). Similarly, the activated platelet-derived growth factor receptor binds both p120-GAP and GRB-1/p85a, a putative GAP for a Rho/Rac-like protein, again linking Ras and Rho/Rac-like proteins. Currently, we can only speculate on the significance of these interactions. A potential pathway from mitogenic receptors to Rac and Rho (including membrane ruffling and stress fibres respectively) has been described[8,9], and these receptors also initiate a pathway that leads to Ras (initiating a cascade of serine/threonine kinases). The pathways are unlikely to be independent, and probably consist of a complex network of GTPase regulators and processors. Identification of structural motifs within these regulators will play an important part in the process of dissecting their functions. □

Mark S. Boguski is at the National Center for Biotechnology Information, National Library of Medicine, NIH, Bethesda, Maryland 20894, USA and Frank McCormick is at Onyx Pharmaceuticals, 3031 Research Drive, Richmond, California 94806, USA.

1. Barbacid, M. *A. Rev. Biochem.* **56,** 779–827 (1987).
2. Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **349,** 117–127 (1991).
3. Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **348,** 125–132 (1990).
4. Gibbs, J. B. & Marshall, M. S. *Microbiol. Rev.* **53,** 171–185 (1989).
5. Bollag, G. & McCormick, F. *A. Rev. Cell Biol.* **7,** 601–632 (1991).
6. Kahn, R. A., Der, C. J. & Bokoch, G. M. *FASEB J.* **6,** 2512–2513 (1992).
7. Fernandez-Sarabia, M. J. & Bischoff, J. R. *Nature* **366,** 274–275 (1993).
8. Ridley, A. J., Paterson, H. F., Johnston, C. L., Diekmann, D. & Hall, A. *Cell* **70,** 401–410 (1992).
9. Ridley, A. J. & Hall, A. *Cell* **70,** 389–399 (1992).
10. Segal, A. W. & Abo, A. *Trends biochem. Sci.* **18,** 43–47 (1993).
11. Pryer, N. K., Wuestehube, L. J. & Schekman, R. *A. Rev. Biochem.* **61,** 471–516 (1992).
12. Roberge, M. *Trends Cell Biol.* **2,** 277–281 (1992).
13. Balch, W. E., Kahn, R. A. & Schwaninger, R. *J. biol. Chem.* **267,** 13053–13061 (1992).
14. Donaldson, J. G., Cassel, D., Kahn, R. A. & Klausner, R. D. *Proc. natn. Acad. Sci. U.S.A.* **89,** 6408–6412 (1992).
15. Lenhard, J. M., Kahn, R. A. & Stahl, P. D. *J. biol. Chem.* **267,** 13047–13052 (1992).
16. Zeuzem, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **89,** 6619–6623 (1992).
17. Moore, M. S. & Blobel, G. *Nature* **1,** 661–663 (1993).
18. Stamnes, M. A. & Rothman, J. E. *Cell* **73,** 999–1005 (1993).
19. Lai, C.-C., Broek, D. & Powers, S. *Molec. cell. Biol.* **13,** 1345–1352 (1993).
20. Mistou, M.-Y. *et al. EMBO J.* **11,** 2391–2397 (1992).
21. Munder, T. & Furst, P. *Mol. cell. Biol.* **12,** 2091–2099 (1992).
22. Mizuno, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **88,** 6442–6446 (1991).
23. Verrotti, A. C. *et al. EMBO J.* **11,** 2855–2862 (1992).
24. Willumsen, B. M. *et al. Mol. cell. Biol.* **11,** 6026–6033 (1991).
25. Li, N. *et al. Nature* **363,** 85–88 (1993).
26. Rozakis-Adcock, M., Fernley, R., Wade, J., Pawson, T. & Bowtell, D. *Nature* **363,** 83–85 (1993).
27. Olivier, J. P. *et al. Cell* **73,** 179–191 (1993).
28. Baltensperger, K. *et al. Science* **260,** 1950–1952 (1993).
29. Skolnik, E. Y. *et al. Science* **260,** 1953–1955 (1993).
30. Botwell, D., Fu, P., Simon, M. & Senior, P. *Proc. natn. Acad. Sci. U.S.A.* **89,** 6511–6515 (1992).
31. Petitjean, A., Hilger, F. & Tatchell, K. *Genetics* **124,** 797–806 (1990).
32. Albright, C. F., Giddings, B. W., Liu, J., Vito, M. & Weinberg, R. A. *EMBO J.* **12,** 339–347 (1993).
33. Drubin, D. G. *Cell* **65,** 1093–1096 (1991).
34. Park, H. O., Chant, J. & Herskowitz, I. *Nature* **365,** 269–374 (1993).
35. Powers, S., Gonzales, E., Christensen, T., Cubert, J. & Broek, D. *Cell* **65,** 1225–1231 (1991).
36. Zhang, X.-f *et al. Nature* **364,** 308–313 (1993).
37. van Aelst, L., Barr, M., Marcus, S., Polverino, A. & Wigler, M. *Proc. natn. Acad. Sci. U.S.A.* **90,** 62213–6217 (1993).
38. Warne, P. H., Viciana, P. R. & Downward, J. *Nature* **364,** 352–355 (1993).
39. Vojtek, A. B., Hollenberg, S. M. & Cooper, J. A. *Cell* **74,** 205–214 (1993).
40. Ron, D. *et al. New Biologist* **3,** 372–379 (1991).
41. Hart, M. J., Eva, A., Evans, T., Aaronson, S. A. & Cerione, R. A. *Nature* **354,** 311–314 (1991).
42. Miki, T., Smith, C. L., Long, J. E., Eva, A. & Fleming, T. P. *Nature* **362,** 462–465 (1993).
43. Tyers, M. *et al. Nature* **333,** 470–473 (1988).
44. Tyers, M., Haslam, R. J., Rachubinski, R. A. & Harley, C. B. *J. cell. Biochem.* **40,** 133–145 (1989).
45. Musacchio, A., Gibson, T., Rice, P., Thompson, J. & Saraste, M. *Trends biochem. Sci.* **18,** 343–348 (1993).
46. Ohtsubo, M. *et al. Genes Dev.* **1,** 585–593 (1987).
47. Bischoff, F. R. & Ponstingl, H. *Proc. natn. Acad. Sci. U.S.A.* **88,** 10830–10834 (1991).
48. Ando, S. *et al. J. biol. Chem.* **267,** 25709–25713 (1992).
49. Uchida, S. *et al. Molec. cell. Biol.* **10,** 577–584 (1990).
50. Burstein, E. S. & Macara, I. G. *Proc. natn. Acad. Sci. U.S.A.* **89,** 1154–1158 (1992).
51. Moya, M., Robert, D. & Novick, P. *Nature* **361,** 460–463 (1993).
52. Burton, J., Roberts, D., Montaldi, M., Novick, P. & De, C. P. *Nature* **361,** 464–467 (1993).
53. Kaibuchi, K. *et al. Molec. cell. Biol.* **11,** 2873–2880 (1991).
54. Hiraoka, K. *et al. Biochem. biophys. Res. Commun.* **182,** 921–930 (1992).
55. Mizoguchi, A., Kim, S., Ueda, T. & Takai, Y. *Biochem. biophys. Res. Commun.* **162,** 1438–1445 (1989).
56. Zahner, J. E. & Cheney, C. M. *Molec. cell. Biol.* **13,** 217–227 (1992).
57. Waldherr, M., Ragnini, A., Schweyen, R. J. & Boguski, M. S. *Nature Genet.* **3,** 193–194 (1993).
58. Seabra, M. C., Brown, M. S., Slaughter, C. A., Sudhof, T. C. & Goldstein, J. L. *Cell* **70,** 1049–1057 (1992).
59. Ohga, N., Kikuchi, A., Ueda, T., Yamamoto, J. & Takai, Y. *Biochem. biophys. Res. Commun.* **163,** 1523–1533 (1989).
60. Settleman, J., Narasimhan, V., Foster, L. C. & Weinberg, R. A. *Cell* **69,** 539–549 (1992).
61. Pawson, T. *Oncogene* **3,** 491–495 (1993).
62. Hancock, J. F. & Hall, A. *EMBO J.* **12,** 1915–1921 (1993).
63. Hart, M. J. *et al. Science* **258,** 812–815 (1992).
64. Chuang, T. H., Xu, X., Knaus, U. G., Hart, M. J. & Bokoch, G. M. *J. biol. Chem.* **268,** 775–778 (1993).
65. Trahey, M. & McCormick, F. *Science* **238,** 542–545 (1987).
66. Hall, A. *Cell* **61,** 921–923 (1990).
67. Koch, C. A., Anderson, D., Moran, M. F., Ellis, C. & Pawson, T. *Science* **252,** 668–674 (1991).
68. Clark, J. D. *et al. Cell* **65,** 1043–1051 (1991).
69. Gutmann, D. H. *et al. Oncogene* **8,** 761–769 (1993).
70. Gideon, P. *et al. Molec. cell. Biol.* **12,** 2050–2056 (1992).
71. Martin, G. A. *et al. Science* **255,** 192–194 (1992).
72. Duchesne, M. *et al. Science* **259,** 525–528 (1993).
73. Wigler, M. *Nature* **346,** 696–697 (1990).
74. Boguski, M. S., Hardison, R. C., Schwartz, S. & Miller, W. *New Biologist* **4,** 247–260 (1992).
75. Marchuk, D. A. *et al. Genomics* **11,** 931–940 (1991).
76. Banu, T. N. *et al. Nature* **356,** 713–715 (1992).
77. DeClue, J. E. *et al. Cell* **69,** 265–273 (1992).
78. The, I., *et al. Nature Genet.* **3,** 62–66 (1993).
79. Li, X. *et al. Cell* **69,** 275–281 (1992).
80. Shannon, K. M. *et al. New Engl. J. Med.* (in the press).
81. Gaul, U., Mardon, G. & Rubin, G. M. *Cell* **68,** 1007–1019 (1992).
82. Wang, Y., Boguski, M., Riggs, M., Rodgers, L. & Wigler, M. *Cell. Regul.* **2,** 453–465 (1991).
83. Kim, J. H. & Powers, S. *Molec. cell. Biol.* **11,** 3894–3904 (1991).
84. Rubinfeld, B. *et al. Molec. cell. Biol.* **12,** 4634–4642 (1992).

85. Hall, C. *et al. J. molec. Biol.* **211,** 11–16 (1990).
86. Otsu, M. *et al. cell* **65,** 91–104 (1991).
87. Skolnik, E. Y. *et al. Cell* **65,** 83–90 (1991).
88. George, J. M. & Clayton, D. F. *Molec. Brain Res.* **12,** 323–329 (1992).
89. LeClerc, S. *et al. J. biol. Chem.* **266,** 17333–17340 (1991).
90. Agnel, M., Roder, L., Vola, C. & Griffin-Shea, R. *Molec. cell. Biol.* **12,** 5111–5122 (1992).
91. Tan, T., Vollmer, P. & Gallwitz, D. *FEBS Lett.* **291,** 322–326 (1991).
92. Walsorth, N. C., Brennwald, P., Kabcenell, A. K., Garrett, M. & Novick, P. *Molec. cell. Biol.* **12,** 2017–2028 (1992).
93. Strom, M., Vollmer, P., Tan, T. J. & Gallwitz, D. *Nature* **361,** 736–739 (1993).
94. Clarke, S. *A. Rev. Biochem.* **61,** 355–386 (1992).
95. Schafer, W. R. & Rine, J. *A. Rev. Genet.* **30,** 209–237 (1992).
96. Boguski, M. S. *J. Lipid Res.* **33,** 957–974 (1992).
97. Boguski, M. S., Murray, A. W. & Powers, S. *New Biologist* **4,** 408–411 (1992).
98. Armstrong, S. A., Seabra, M. C., Sudhof, T. C., Goldstein, J. L. & Brown, M. S. *J. biol. Chem.* **268,** 12221–12229 (1993).
99. Andres, D. A., Goldstein, J. L., Ho, Y. K. & Brown, M. S. *J. biol. Chem.* **268,** 1383–1390 (1993).
100. Reiss, Y., Brown, M. S. & Goldstein, J. L. *J. biol. Chem.* **267,** 6403–6408 (1992).
101. Reiss, Y. *et al. J. biol. Chem.* **266,** 10672–10677 (1991).
102. He, B. *et al. Proc. natn. Acad. Sci. U.S.A.* **88,** 11373–11377 (1991).
103. Gomez, R. *et al. Biochem. J.* 25–31 (1993).
104. Kohl, N. E. *et al. Science* **260,** 1934–1937 (1993).
105. James, G. L. *et al. Science* **260,** 1937–1942 (1993).
106. Mayer, M. L., Caplin, B. E. & Marshall, M. S. *J. biol. Chem.* **267,** 20589–20593 (1993).
107. Zhang, F. *J. biol. Chem.* (in the press).
108. Ohya, Y. *et al. J. biol. Chem.* **266,** 12356–12360 (1991).
109. Seabra, M. C., Goldstein, J. L., Sudhof, T. C. & Brown, M. S. *J. biol. Chem.* **267,** 14497–14503 (1992).
110. Andres, D. A. *et al. Cell* **73,** 1091–1099 (1993).
111. Armstrong, S. A., Seabra, M. C., Sudhof, T. C., Goldstein, J. L. & Brown, M. S. *J. biol. Chem.* **268,** 12221–12229 (1993).
112. Rossi, G., Jiang, Y., Newman, A. P. & Ferro-Novick, S. *Nature* **351,** 158–161 (1991).
113. Jiang, Y., Rossi, G. & Ferro-Novick, S. *Nature* **366,** 84–86 (1993).
114. Li, R., Havel, C., Watson, J. A. & Murray, A. W. *Nature* **366,** 82–84 (1993).
115. Settleman, J., Albright, C. F., Foster, L. C. & Weinberg, R. A. *Nature* **359,** 153–154 (1992).
116. McGlade, J. *et al. EMBO J.* **12,** 3073–3081 (1993).
117. Shou, C., Farnsworth, C. L., Neel, B. G. & Feig, L. A. *Nature* **358,** 351–354 (1992).
118. Wei, W. *et al. Proc. natn. Acad. Sci. U.S.A.* **89,** 7100–7104 (1992).
119. Martegani, E. *et al. EMBO J.* **11,** 2151–2157 (1992).
120. Simon, M. A., Bowtell, D. D. L., Dodson, G. S., Laverty, T. R. & Rubin, G. M. *Cell* **67,** 701–716 (1991).
121. Bairoch, A. *Nucleic Acids Res.* **19,** 2241–2245 (1991).
122. Bairoch, A. & Boeckmann, B. *Nucleic Acids Res.* **20,** 2019–2022 (1992).
123. Kohl, N. E. *et al. J. biol. Chem.* **266,** 18884–18888 (1991).
124. Boguski, M. S., Love, T. M. J. & Tolstoshev, C. M. *Nature Genet.* **4,** 332–333 (1993).
125. Schuler, G. D., Altschul, S. F. & Lipman, D. J. *Proteins struct. funct. Genet.* **9,** 180–190 (1991).
126. Boguski, M. S., Bairoch, A., Attwood, T. K. & Michaels, G. S. *Nature* **358,** 113 (1992).
127. Hart, M. J. *et al. J. biol. Chem.* (in the press).
128. Gulbins, E. *et al. Science* **260,** 822–825 (1993).

ACKNOWLEDGEMENTS. We thank J. Goldstein, P. Polakis, S. Powers, M. Ruggieri, J. Hancock, G. Bollag, A. Bernards, J. Mulligan, A. Murray, S. Ferro-Novick and P. Casey for helpful discussions and providing data before publication.

ARTICLES

The 2.2 Å crystal structure of transducin-α complexed with GTPγS

Joseph P. Noel*, Heidi E. Hamm†‡ & Paul B. Sigler*§

* Department of Molecular Biophysics and Biochemistry and the Howard Hughes Medical Institute, Yale University, 295 Congress Avenue, Boyer Center for Molecular Medicine, Room 154, New Haven, Connecticut 06510, USA
† Department of Physiology and Biophysics, University of Illinois at Chicago, Chicago, Illinois 60680, USA
‡ Istituto di Fisiologia Generale e Chimica Biologica, Universita' di Sassari, 07100 Sassari, Italy

The 2.2 Å crystal structure of activated rod transducin, $G_{t\alpha}$·GTPγS, shows the bound GTPγS molecule occluded deep in a cleft between a domain structurally homologous to small GTPases and a helical domain unique to heterotrimeric G proteins. The structure, when combined with biochemical and genetic studies, suggests: how an activated receptor might open this cleft to allow nucleotide exchange; a mechanism for GTP-induced changes in effector and receptor binding surfaces; and a mechanism for GTPase activity not evident from previous data.

HERE we describe the three-dimensional structure of the activated α-subunit of a heterotrimeric ($\alpha\beta\gamma$) G protein, namely the complex of transducin-α with the GTP analogue GTPγS. The model and the mechanisms inferred from it were derived from an X-ray crystal structure refined to a resolution of 2.2 Å. The model provides a stereochemical context for the extensive biochemical and mutagenic studies[1] on the highly homologous family of G-protein-coupled signal transduction systems.

Each heterotrimeric G protein is functionally coupled to a specific seven-helix transmembrane receptor, but all share a common mechanism by which information is transduced from the receptor through a GTP-activated α-subunit or a free $\beta\gamma$ heterodimer to specific downstream effectors. GTP's γ-phosphate acts as a molecular switch that relays information in the form of a conformational change. This conformational change leads to a decreased affinity for both the receptor and the $\beta\gamma$ heterodimer, and an increased specific affinity for and activation of an effector enzyme or channel which ultimately leads to a change in the cellular concentration of a second messenger such as a cyclic nucleotide, Ca^{2+}, inositol phosphates, or diacylglycerol. The released $\beta\gamma$ subunit can itself lead to the activation or modulation of some effectors[2,3]. A GTPase-controlled timing mechanism inherent in all α-subunits and, in some cases, modulated by other proteins[4,5] returns the GTP-activated α-subunit to the inactive GDP-bound heterotrimer.

In the vertebrate light response, inactive heterotrimeric transducin ($G_{t\alpha\beta\gamma}$·GDP) binds the photoactivated rhodopsin (metarhodopsin II) and releases GDP. GTP binds firmly to the α-subunit of the receptor-bound transducin causing $G_{t\alpha}$·GTP to separate from both the receptor and the $\beta\gamma$ heterodimer. The $G_{t\alpha}$·GTP complex activates its effector, cGMP-specific phosphodiesterase ($\alpha\beta\gamma_2$ cGMP–PDE) by binding to its inhibitory γ-subunit(s).

Activated $\alpha\beta$ cGMP–PDE lowers the concentration of

§ To whom correspondence should be addressed.

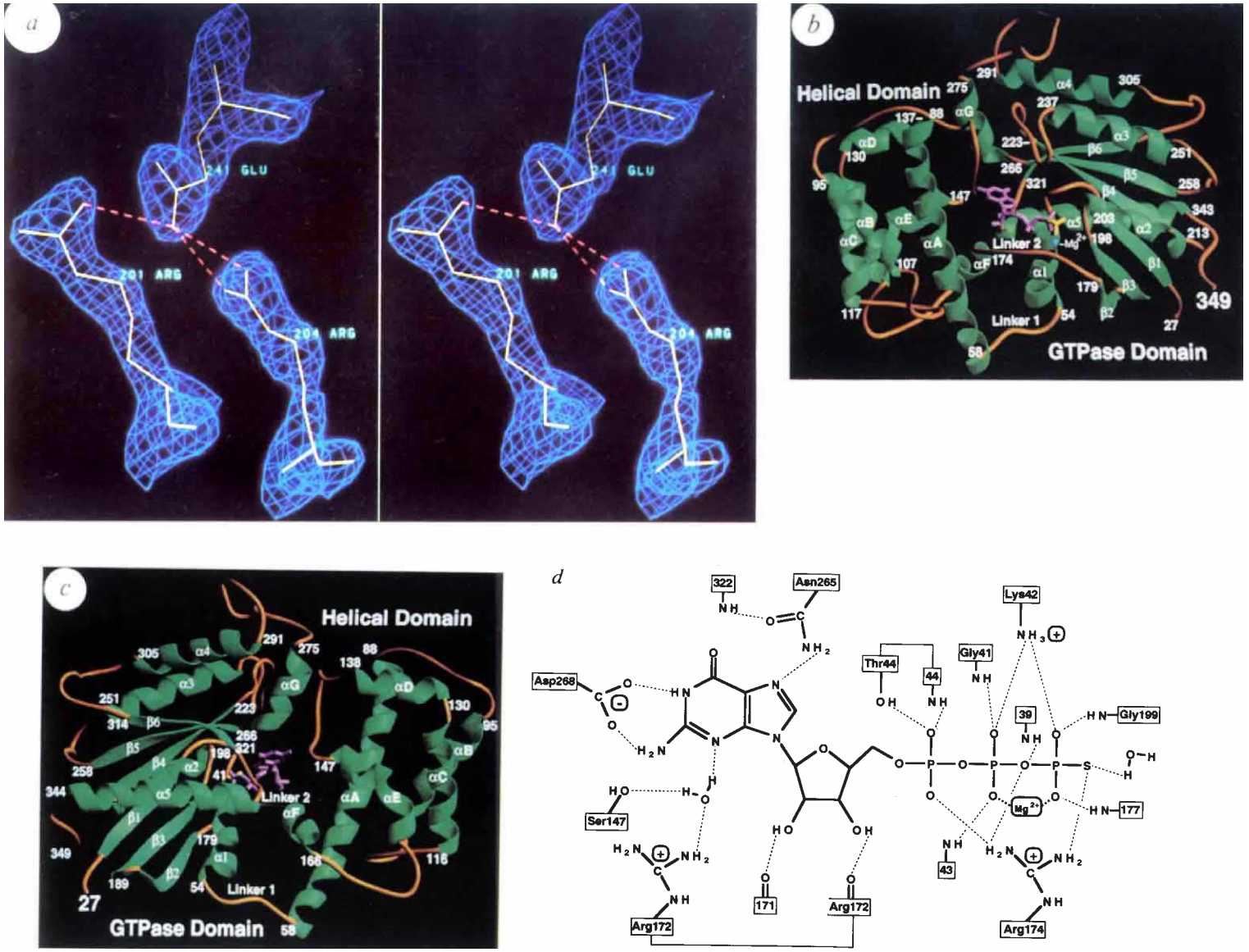


FIG. 1 The overall structure of (-25)$G_{t\alpha}$·GTPγS. *a*, Representative electron density. The map was calculated at 2.2 Å resolution with $(2F_o - F_c)$ amplitudes and model phases and is contoured at 2.0σ. This example illustrates electrostatic linkages that convey the signal to the effector binding region (see Fig. 4*a*, *b*, *c* for details.) *b*, Ribbon drawing (Ribbons[43]). The γ-phosphate appears yellow while the rest of the nucleotide is lavender. The catalytically essential Mg^{2+} is represented as a cyan sphere. For clarity, labels for the secondary structural elements of the core GTPase domain reflect the convention used to describe Ras[17]. The seven additional helices which are unique in the $G_{t\alpha}$·GTPγS structure are labelled A through G. Although the crystallized construct spans residues 26 through 350, the crystallographically observable N and C termini are Ala 27 and Leu 349. *c*, Alternative view of the overall fold of the complex rotated 170° around a vertical axis from the view in (b). *d*, Schematic illustration of the major interactions securing GTPγS in the nucleotide binding cleft. Hydrogen bonds are denoted by dashed lines. Only the interactions observed in all three complexes are shown. All of the figures were constructed from the coordinates of one of the complexes, but they are representative of the other two. (The r.m.s. deviation of *all* atoms for protomer pairs is 1.14 Å for protomer A to B, 0.95 Å for protomer A to C and 1.03 Å for protomer B to C. The r.m.s. deviation for backbone atoms only is 0.50 Å for A to B, 0.42 Å for A to C and 0.44 Å for B to C.)

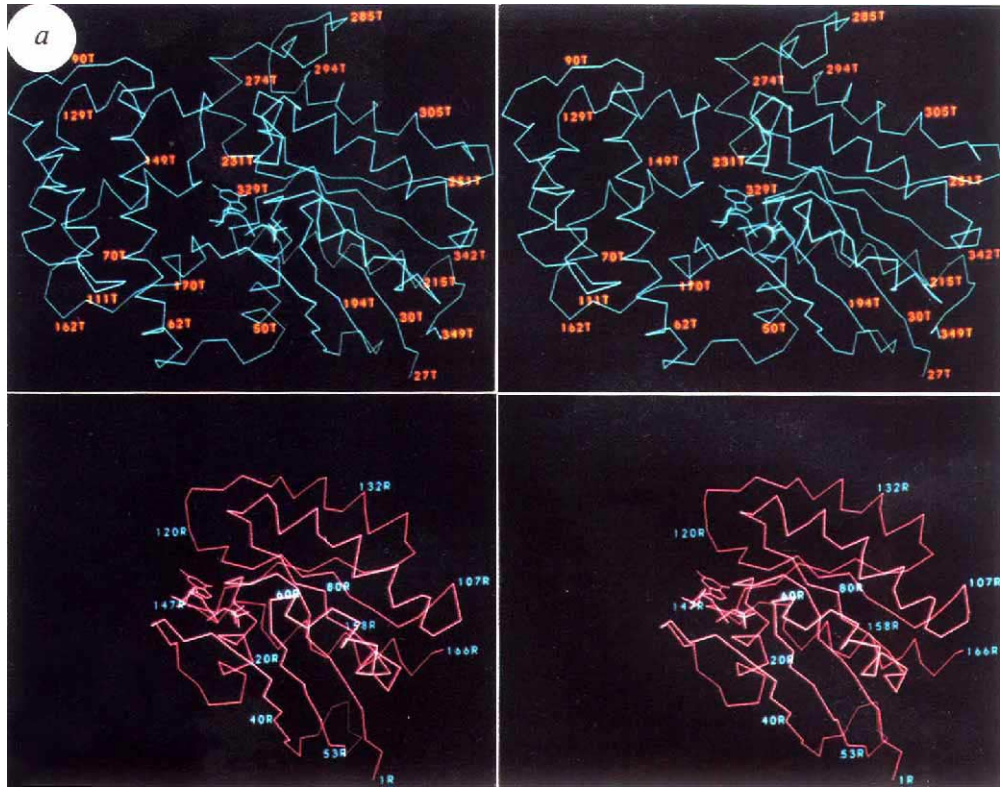


FIG. 2 Primary and tertiary structural alignment of (-25)$G_{t\alpha}$·GTPγS with representative GTPases. *a*, Stereoview of the α carbon trace and bound nucleotides for $G_{t\alpha}$·GTPγS and Ras·GMP.PNP[12] in the same orientation. $G_{t\alpha}$·GTPγS appears in light blue and residue numbers are appended with T. Ras·GMP.PNP appears in pink and its sequence numbers are appended with R. The root mean square difference in backone positions (excluding surface loops) between the GTPase domains of $G_{t\alpha}$ and Ras–GMP.PNP[12] is 1.85 Å. *b*, Primary sequence alignment. The observed secondary structural elements as deduced from $G_{t\alpha}$ are indicated below the sequences. The location of functional domains are indicated above the sequences as are the sequence numbers for $G_{t\alpha}$. Regions considered as 'inserts' in the GTPase domain are included above the aligned sequences. The amino acids are donoted by their one letter codes. Gaps in the aligned sequences are indicated by dots. Sources of sequence information: bovine $G_{s\alpha}$[44], bovine $G_{i\alpha}$[45]. bovine $G_{o\alpha}$[46], mouse $G_{q\alpha}$[47], bovine $G_{z\alpha}$[48], bovine rod $G_{t\alpha}$[49], Ras[50], *E. coli* EF-Tu[51].

cGMP, which closes cGMP-gated cation channels and hyperpolarizes the retinal rod cell, thereby generating the nerve impulse. $G_{t\alpha}$·GDP releases PDE-γ leading to a reformation of both the inactive cGMP–PDE $\alpha\beta\gamma_2$ complex and the inactive $G_{t\alpha\beta\gamma}$·GDP heterotrimer which is primed for reactivation by metarhodopsin II (refs 6–8).

Structure determination and refinement

A 325-amino-acid fragment of bovine retinal rod cell $G_{t\alpha}$·GTPγS lacking residues 1–25 was obtained by proteolysis of the full-length GTPγS complex with endoproteinase lysC (ref. 9). Although the removal of the N terminus results in a loss of $\beta\gamma$ binding, the GTPγS complex fully retains its capacity to activate its target enzyme cGMP–PDE. The best crystals (space group $P2_1$) were grown at −12.5 °C in high concentrations of cacodylate buffer and $CaCl_2$ and contained three complexes per asymmetric unit. Phases from two heavy-atom derivatives of marginal quality were substantially improved by a solvent-flattening procedure employing maximum entropy routines (Z. Otwinowski) and subsequent averaging of the non-crystallographically related density in the asymmetric unit (Table 1).

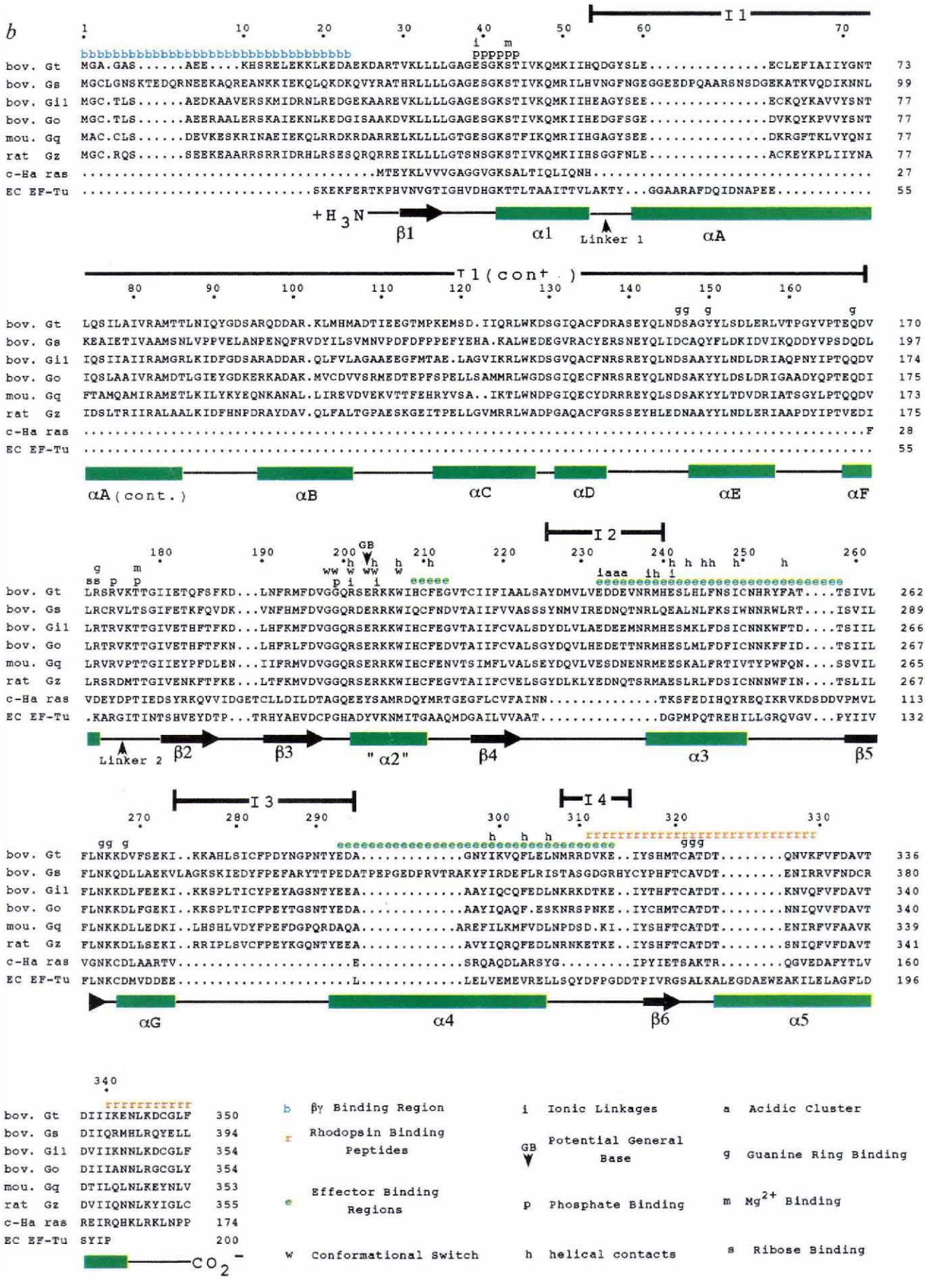
A model for each of the three non-crystallographically related complexes was built into the final averaged map with the program FRODO (ref. 10) and refined with XPLOR 3.1 (ref. 11) to an *R*-factor of 19.0% for all data, and 17.8% for data greater than 2σ between 6.0 and 2.2 Å resolution. The free *R*-factor for a randomly omitted subset (10%) of data (no cut-off) was 30.4 and 28.7% for data greater than 2.0σ. The average deviation from standard values is 0.011 Å in bond distance and 1.8° in bond angle; the Φ–Ψ plot showed all residues except for Asn 54 and some glycines within 10° of the allowed regions. A representative segmet of the unaveraged electron density map is shown in Fig. 1*a*.

The domain architecture

The (-25)$G_{t\alpha}$·GTPγS complex consists of two easily delineated domains (Fig. 1*b* and *c*) flanking a deep guanine-nucleotide-binding cleft (Fig. 1*d*). One domain is a highly conserved GTPase domain that is structurally similar to the $\alpha\beta$ fold of the c-H-*ras* protein, $p21^{Ha\text{-}ras}$ or Ras[12] (Fig. 2*a*), even though the sequence indentity is low (19.0%; Fig. 2*b*). This domain, also seen in the protein synthesis elongation factor EF-Tu (ref. 13), consists of a six-stranded β-sheet ($\beta1$–$\beta6$) surrounded by a set of five helices ($\alpha1$–$\alpha5$). The other domain is a highly helical domain unique to heterotrimeric G proteins. Altogether, there are four dfferences between the (-25)$G_{t\alpha}$·GTPγS complex and a consensus GTPase fold as represented by Ras: one is the large helical domain just mentioned (I1); the other three occur as sequence insertions within loops of the GTPase domain (I2, I3, I4).

The highly helical domain (I1) spans residues 59–172 (Fig. 2*b*). It is connected to the GTPase domain by two polypeptide segments, hereafter referred to as linker 1 and linker 2. Linker 1, spanning residues 54–58, follows $\alpha1$. Linker 2, spanning residues 173–179, immediately precedes $\beta2$. The helical domain and the GTPase domain clamp down upon and completely bury the GTPγS molecule and its associated Mg^{2+} ion. EF-Tu (refs 13–15) and the smaller GTPase, Ras[12], which both lack this helical domain, bind the nucleotide in a partially exposed surface cleft. The inaccessibility of GTPγS in $G_{t\alpha}$ suggests that in order to exchange guanine nucleotides there must be a conformational change that opens the cleft. The helical domain encompasses a central α-helical scaffold of 28 residues which supports five smaller helices through hydrophobic interactions. The seemingly rigid internal structure of the helical domain suggests that conformational changes in the linkers would be amplified into 'en bloc' movements of the helical domain, making it a lid on the nucleotide binding site.

Insert I2 residues between $\beta4$ and $\alpha3$ and makes critical ionic contacts to both the conserved phosphate-binding loop[16] and the so-called switch II region[17], residues 198–215. We will invoke these interactions when we consider a mechanism for the propagation of a conformational change from the γ-phosphate binding region to the effector (PDE-γ) binding loops. The third


b
I 1
bov. Gt
bov. Gs
bov. Gi1
bov. Go
mou. Gq
rat Gz
c-Ha ras
EC EF-Tu
+H3N
β1
α1
Linker 1
αA
T1 (cont.)
αA (cont.)
αB
αC
αD
αE
αF
I 2
GB
Linker 2
β2
β3
"α2"
β4
α3
β5
I 3
I 4
αG
α4
β6
α5
CO2−
b βγ Binding Region
r Rhodopsin Binding Peptides
e Effector Binding Regions
w Conformational Switch
i Ionic Linkages
GB Potential General Base
p Phosphate Binding
h helical contacts
a Acidic Cluster
g Guanine Ring Binding
m Mg2+ Binding
s Ribose Binding

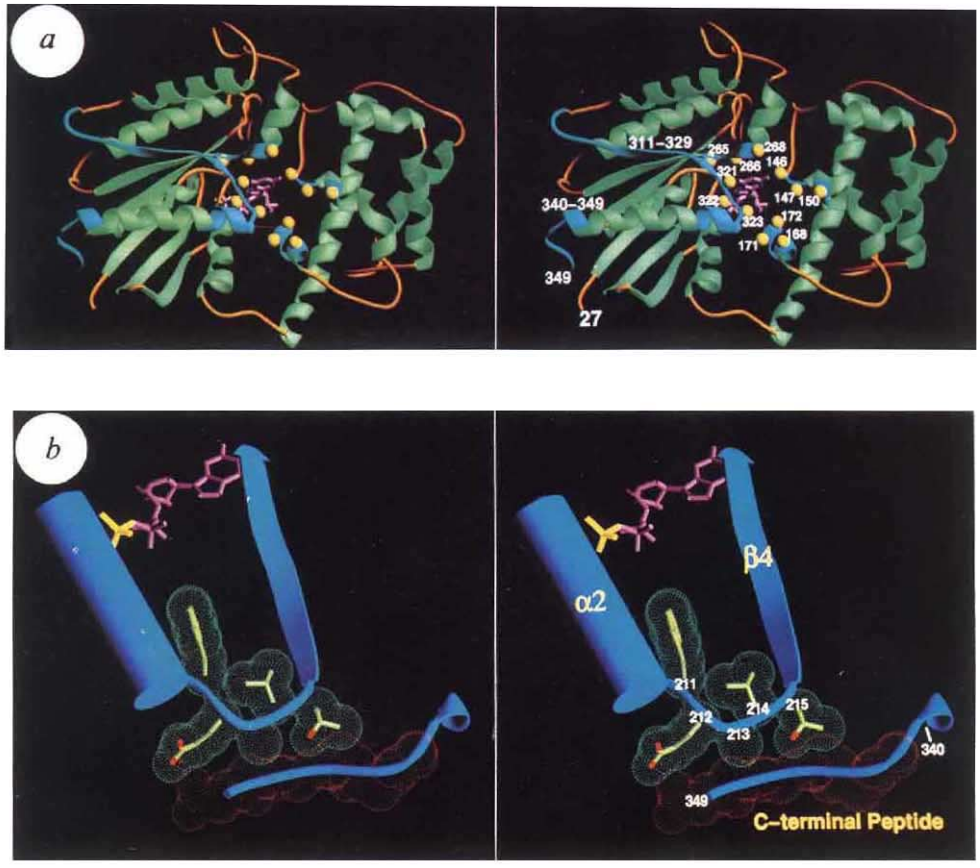


FIG. 3 Linkage between nucleoside binding sites and the receptor interaction surface. *a*, Stereoview of the proposed receptor binding face of $G_{t\alpha}$. Regions involved in receptor binding, 311–329 and 340–350 (ref. 19), and segments participating in nucleoside binding are coloured in light blue. The yellow spheres denote relevant contacts, direct or supporting, to the nucleoside. *b*, Stereoview of the van der Waals contact surface formed between the GTP regulated $\alpha 2/\beta 4$ loop side chains and the backbone of the rhodopsin binding C-terminal peptide. The C-terminal backbone depicted (light blue) spans residues 340–349 while the van der Waals surface of this backbone (red dot surface) encompasses residues 343 to 349. *c*, Stereoview of the cluster of residues highlighted as yellow spheres in *a*. Cys 321 fixes the side chain of Thr 323 which forms a favourable van der Waals surface with the guanine ring. Ala 322's backbone amide positions the side chain of Asn 265 which, in turn, donates a hydrogen bond to the N7 nitrogen of the base. Asp 268 accepts hydrogen bonds from the N1 and N2 nitrogens of guanine. Lys 266 provides a second van der ►

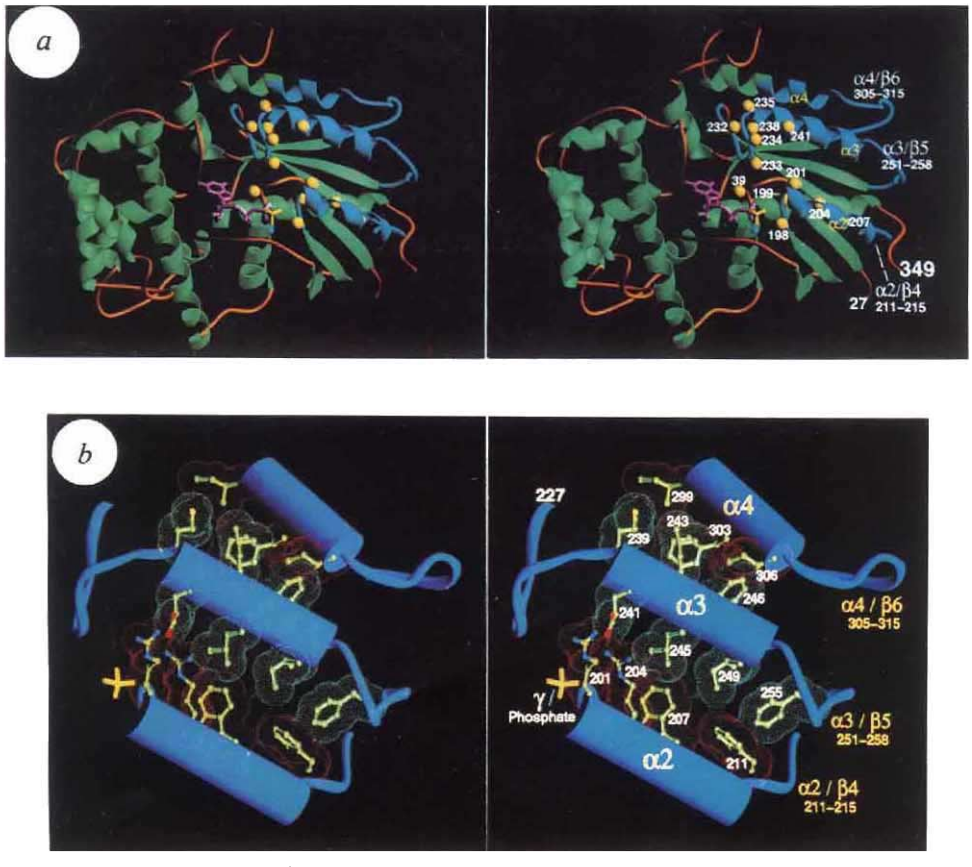
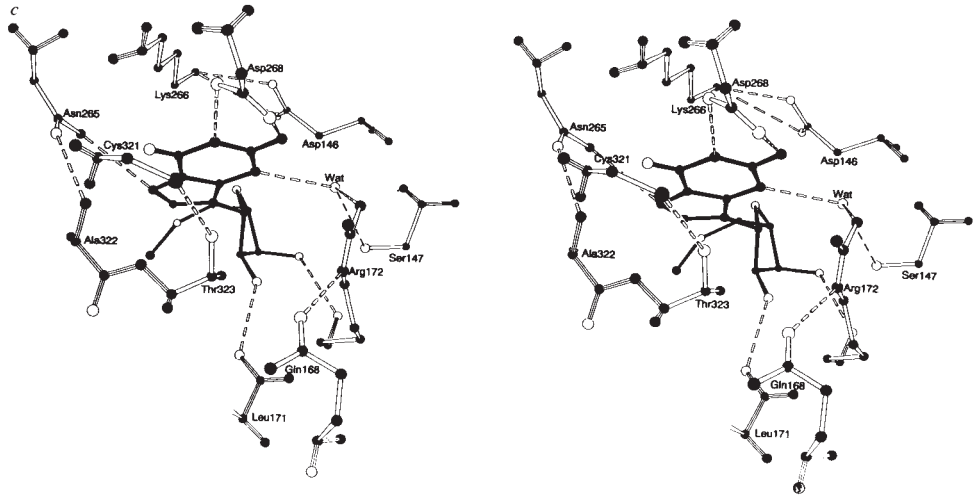


FIG. 4 Linkage between γ-phosphate binding and effector interaction surfaces. *a*, Stereoview highlighting in light blue regions involved in the functional coupling between GTP binding (peptide Gly 198/Gly 199) and the effector binding surfaces ($\alpha 2/\beta 4$, $\alpha 3/\beta 5$, $\alpha 4/\beta 6$) mediated through $\alpha 2$, $\alpha 3$ and $\alpha 4$. In addition, a fourth region comprising an acidic cluster of residues, Glu 232, Asp 233, Asp 234 and Glu 235 is shown and comprises a fourth effector binding surface (J. Mills *et al.*, in preparation). In addition to the acidic cluster, the yellow spheres emphasize specific contacts between the ►



Waals surface for the guanine ring and is locked in place by the side chain carboxylate of Asp 146 of the helical domain. Both the 2′ and 3′ ribose hydroxyls donate hydrogen bonds to the backbone carbonyls of residues 171 and 172. The N3 of the guanine ring participates in a water mediated contact fixed in place by the side chains of Ser 147 and Arg 172. Gln 168 and Tyr 150 (not shown) buttress the side chain of Arg 172 through hydrogen bonds and van der Waals forces, respectively. The nucleoside bonds are filled and the protein's backbone bonds are partially filled. Hydrogen bonds are represented by broken bonds. C atoms are black, N, P and S atoms are partially filled and O atoms are white.

major insert, I3, is situated between $\beta5$ and $\alpha4$ and encompasses αG. The fourth, I4, is a small insert from residues 309 to 314 connecting $\alpha4$ and $\beta6$ and plays a key role in effector recognition.

Guanine-nucleotide binding

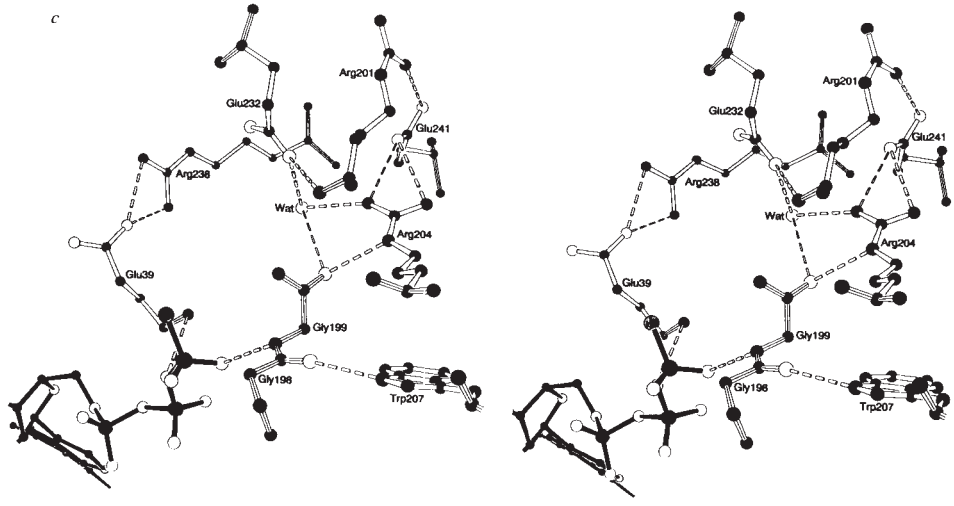
Triphosphate binding. The triphosphate portion of GTPγS coordinates an essential Mg^{2+} ion and accepts hydrogen bonds from the backbone amides of Glu 39, Gly 41, Lys 42, Ser 43, Thr 177 and Gly 199, and the side chains of Lys 42 and Arg 174. The Mg^{2+} ion is coordinated in an octahedral arrangement in which two ligands are non-bridging oxygens of the β- and γ-phosphates, two ligands are the side-chain hydroxyl groups of Ser 43 and Thr 177, and two opposing ligands are water molecules that are secured by hydrogen bonds to the side chains of Ser 43, Thr 177 and Asp 196, and a non-bridging oxygen on the α-phosphate.

The binding of the backbone amides of Ser 43 and Thr 177 to the β- and γ-phosphates positions the side chains to interact with the Mg^{2+} ion and with its associated water ligands. The side chain of Arg 174, which is the site for GTPase inhibiting ADP-ribosylation by cholera toxin[18], is hydrogen-bonded to the γ-thiophosphate's sulphur, the bridging oxygen between the β- and γ-phosphate, and a non-bridging oxygen on the α-phosphate (Fig. 1*d*).

Nucleoside binding. The 2′ and 3′ hydroxyls of the sugar donate hydrogen bonds to the backbone carbonyls of Leu 171 and Arg 172 in αF. The guanine ring is bound by an interaction that is unique to the $G_{t\alpha}$ structure, in which the N3 ring nitrogen forms water-mediated interactions with the side chains of Ser 147 and Arg 172. It is noteworthy that the nucleoside contacts (Ser 147, Leu 171 and Arg 172) and the phosphate contacts (Arg 174 and Thr 177) arise from either the helical domain or linker 2, which, as will be discussed in a later section, has important implications for receptor-regulated nucleotide exchange. In addition to these guanine interactions that are unique to G proteins, a series of interactions conserved in the GTPase superfamily further stabilizes the guanine base through hydrogen bonds from side chains of Asn 265 to the N7 ring nitrogen and Asp 268 to the N1 ring nitrogen and exocyclic 2-amine (Fig. 1*d*). Finally, the guanine ring is sandwiched by van der Waals contacts between the methylene carbons of Lys 266 and the methyl group of Thr 323. Lys 266 is well stabilized by a salt bridge to Asp 146 of the helical domain, and Thr 323 is also locked in place by a supporting hydrogen bond with the side chain of Cys 321. A mechanistically important feature of this system is the elegant manner in which interactions with one portion of the nucleotide support contacts with another. It is likely that these tightly coupled interactions potentiate a highly cooperative receptor-mediated disassembly of the elements that so strongly secure GDP and GTP in the nucleotide-binding cleft.

Rhodopsin binding and nucleotide exchange

The C-terminal receptor-binding site. A peptide designed to mimic the C-terminal 11 residues of $G_{t\alpha}$ not only competes with $G_{t\alpha\beta\gamma}$ binding to photoactivated rhodopsin, but stabilizes rho-



phosphate binding loop (Glu 39), the $\alpha2$ switch region (Gly 198–Trp 207), and the acidic loop (Glu 232–Glu 241). *b*, Stereoview of the $\alpha2/\alpha3/\alpha4$ interface. The dot surface (van der Waals) emphasizes the stereochemical coupling between the conformationally regulated $\alpha2$ helix and $\alpha3$ and $\alpha4$. *c*, Detailed stereochemical view of the coupling between the γ-phosphate binding region (Glu 39, Gly 198 and Gly 199), the conformationally regulated $\alpha2$ helix (Arg 201, Arg 204 and Trp 207) and the I2 loop (Glu 232, Arg 238 and Glu 241).

TABLE 1 Experimental data on crystal structure determination and refinement

			Quality of MIR phases						
Resolution limit (Å)	12.6	8.4	6.3	5.0	4.2	3.6	3.2	2.8	Total
CH_3-Hg(OOC.CH_3)$_2$									
Number of reflections	280	760	1,478	2,404	3,571	4,893	3,869	—	17,255
Phasing power	2.53	2.29	2.49	2.05	1.33	1.19	1.37	—	1.49
$SmCl_2$									
Number of reflections	242	758	1,477	2,405	3,573	4,946	6,612	8,264	28,277
Phasing power	2.05	1.04	1.03	0.78	0.47	0.36	0.33	0.31	0.49
Mean figure of merit	0.76	0.79	0.74	0.67	0.55	0.45	0.33	0.19	0.41
			Native data and refinement						
Resolution limit (Å)	5.41	3.76	3.17	2.84	2.61	2.44	2.31	2.20	Total
Number of reflections	3,977	7,834	7,853	7,803	7,655	7,507	6,101	4,559	53,289
Predicted	4,019	7,839	7,863	7,835	7,820	7,801	7,763	7,792	58,732
R_{sym}	0.026	0.029	0.037	0.049	0.067	0.094	0.108	0.135	0.039
R-factor (all data)	0.173	0.121	0.157	0.205	0.246	0.277	0.304	0.330	0.190
R-factor (>2.0σ)	0.172	0.121	0.155	0.197	0.231	0.248	0.255	0.268	0.178
Free R-factor (all data)									0.304
Free R-factor (>2.0σ)									0.287

$R_{sym} = \sum_h \langle |I_h' - I_h| \rangle / \sum_h I_h$ where $\langle |I_h - I_h| \rangle$ is the average absolute deviation of reflections I_h from the average $I_{h'}$ of its symmetry and Friedel related equivalents. Phasing power is $\sum |F_h^c| / [\{|F_p^o| \exp(i\Phi_c) + F_h^c| - |F_{ph}^o|\}]$, where F_h^c = heavy atom structure factor and $|F_p^o|$ and $|F_{ph}^o|$ are observed amplitudes for the protein and heavy-atom derivative, respectively, Φ_c is the calculated phase, and the sum is over all observations. Figure of merit is $\int P(\Phi) \exp(i\Phi)\, d\Phi / \int P(\Phi)\, d\Phi$, where P is the probability distribution of Φ, the phase angle. $G_{t\alpha}$·GTPγS was purified from photolysed bovine rod outer segments (ROS) by selective extraction with GTPγS and Mg^{2+} as described[9]. (-25)$G_{t\alpha}$·GTPγS, residues 26–350, was prepared from full-length $G_{t\alpha}$·GTPγS by overnight proteolysis with endoproteinase LysC (Boehringer Mannheim) at 4 °C in 20 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 1 mM dithiothreitol (DTT) and purified on a Mono-Q (Pharmacia) column with a gradient of 0–225 mM NaCl in 20 mM Tris-HCl, pH 7.5, 0.5 mM DTT. The fractions containing $G_{t\alpha}$·GTPγS were concentrated and stored at −80 °C in 5 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 25 mM β-mercaptoethanol, 40% (v/v) glycerol at ~40 mg ml^{-1}. The purified protein produced a single band on an overloaded silver-stained denaturing SDS gel and the N-terminal sequence was verified by electroblotting on to polyvinylidene fluoride (Millipore) membrane and sequenced (K. Stone, HHMI, Yale University). Diffraction quality crystals were grown at −12.5 °C from microseeded hanging drops (3–5 µl) containing 20–25 mg ml^{-1} protein, 166 mM sodium cacodylate buffer, pH 6.0, 223 mM $CaCl_2$, 25 mM β-mercaptoethanol, 20% (v/v) glycerol, 3.33% (w/v) PEG 8000 over reservoirs of 5% (w/v) PEG 8000, 250 mM sodium cacodylate buffer, pH 6.0, 350 mM $CaCl_2$, 25 mM β-mercaptoethanol, 30% (v/v) glycerol. Crystals appeared overnight and grew typically to 0.1 mm × 0.1 mm × 0.2 mm over several weeks. They were stabilized at 4 °C in 15% (w/v) PEG 8000, 50 mM sodium cacodylate buffer, pH 6.0, 25 mM β-mercaptoethanol, 30% (v/v) glycerol. The crystals belong to space group $P2_1$ with a = 74.5 A, b = 108.3 A, c = 79.0 A, β = 112.3°. The Sm^{2+} and CH_3-Hg^+ derivative crystals were equilibrated with the stabilizer containing 10 mM $SmCl_2$ or 0.05 mM methyl mercuric acetate (no β-mercaptoethanol), respectively. Parent data and the Sm^{2+} derivative data were each collected from single quick-frozen crystals (−150 °C) on the RAXIS II, processed with DENZO (Z. Otwinowski), and scaled with SCALEPACK (Z. Otwinowski). The CH_3-Hg^+ derivative data were collected on a Siemens Xentronics area detector, processed with XDS[38,39], and scaled with SCALEPACK. The parent and derivative data sets were placed on the same relative scale with SCALEPACK. Heavy-atom positions were determined initially from a combination of isomorphous and anomalous difference Patterson maps, improved with cross-difference Fourier maps, and refined with ML-PHARE[40]. A total of 21 Sm sites and 12 CH_3-Hg sites were located in the asymmetric unit. The MIR phases were improved sequentially by solvent flattening employing maximum entropy routines (Z. Otwinowski) and non-crystallographic averaging with the program SQUASH[41]. The initial superimposition matrices relating the three (-25)$G_{t\alpha}$·GTPγS's complexes were determined from the positions of the three triphosphates fit from the MIR map and three CH_3-Hg^+ binding sites and refined in SQUASH. The three complexes are related by an approximate 3_1 screw axis. The averaging envelope encompassing one of the complexes was constructed manually by filling a clearly delineated protomer envelope with a set of randomly placed dummy atoms with FRODO[10] and converted into the CCP4 map format with ATOMSTOMASK (J. Friedman). A nearly complete model for each of the three non-crystallographically related complexes was built into the final averaged map with FRODO and refined by simulated annealing with XPLOR 3.1 (ref. 11) starting at 4,000 K (0.3 fs time step). The non-crystallographic symmetry was loosely restrained during the first two rounds of refinement at 2.8 Å and 2.5 Å resolution. Local errors were corrected with the aid of the original experimentally phased maps and model phased (2Fo − Fc) and (Fo − Fc) maps. All three models were then refined at 2.2 Å with tightly restrained neighbouring B-factors (r.m.s. deviations in B-factor between bonded atoms is 1.75 $Å^2$) and no local symmetry restraints. Water sites were gradually and systematically added to the structure with the aid of suitably contoured (Fo − Fc) maps during the later stages of refinement and accepted if they (1) reappeared with strong density in (2Fo − Fc) maps, (2) participated in at least one hydrogen-bonded interaction, and (3) refined to an isotropic B-factor of no more than 70 $Å^2$ with unit occupancy. The free R-factor[42] consistently decreased throughout the refinement. The current model lacks the N-terminal Asp 26 for all three complexes, residues 343–350 in two of the complexes, and Phe 350 in the third. In addition to the polypeptide backbone, the current model contains three GTPγS molecules, three Mg^{2+} ions, six cacodylate-modified cysteines (Cys 61 and Cys 210) modelled as single As atoms and 826 solvent molecules modelled as water.

dopsin's active Meta II photochemical state[19] (50% inhibitory concentration, IC_{50} = 0.04 mM[20]). In addition, work with $G_{i\alpha}$/$G_{q\alpha}$ chimaeras has established a critical role for the C terminus in specifying the α-subunit's binding to its cognate receptor[21]. Finally, antibodies against the C terminus of G_α (ref. 22), and ADP-ribosylation by pertussis toxin of a cysteine side chain[23,24] four residues from the C terminus, uncouple many G proteins from their receptors. In two of the molecules represented in the asymmetric unit, the final eight residues are somewhat disordered. In a third complex, the C-terminal residues 343–349 of $G_{t\alpha}$ can be seen to make van der Waals contacts with the $\alpha 2$/$\beta 4$ loop, residues 212–215 (Fig. 3*a* and *b*). This segment is part of the switch II region (residues 198–215), which in $G_{t\alpha}$ is chemically coupled through Gly 198 and Gly 199 to the terminal phosphate of GTP, and whose conformation is known to undergo a major conformational rearrangement upon binding GTP from altered proteolysis[9,25,26], from fluorescence changes of the $\alpha 2$ residue Trp207 (ref. 27), and from the structures of activated and unactivated Ras[17]. We speculate with appropriate caution that the change in $\alpha 2$ and the $\alpha 2$/$\beta 4$ loop can be propagated to the abutting C-terminal residues 343–349 through direct interactions, thereby coupling the nature of the bound nucleotide to the specific affinity between G_α and its cognate receptor.

Receptor catalysed GDP/GTP exchange. $G_{t\alpha}$, Ras and EF-Tu differ in their affinity for guanine nucleotides. Exchange factors accelerate nucleotide exchange in both EF-Tu[28] and Ras[29], but a significant basal exchange rate persists in their absence. Exchange does not occur for transducin in the absence of activated rhodopsin. The structural feature that distinguishes $G_{t\alpha}$ from both Ras and EF-Tu is the large helical domain that forms

one wall of the nucleotide-binding cleft and thereby occludes the nucleotide. Activated rhodopsin probably opens the nucleotide-binding cleft.

The stereochemical links between the receptor-binding surface and the elements that clamp the nucleotide in the receptor-free heterotrimeric G_αs are highlighted in Fig. 3*a* and *c*. In addition to the previously described C-terminal peptide, a second peptide spanning $G_{t\alpha}$ residues 311–329 has also been shown to bind to photoactivated rhodopsin by competition assays[19]. Residues 320–323 of this peptide contribute directly to nucleoside binding through a van der Waals contact with the guanine ring. This segment is stereochemically linked to the nucleotide-binding cleft as illustrated in Fig. 3*a* and *c*, and presumably modulates nucleotide affinity in a receptor-regulated fashion.

In addition to the conserved nucleoside-binding elements of the GTPase domain, the present structure now shows the unique contributions to nucleotide binding and most likely to receptor-regulated exchange that arise from the helical domain (Fig. 3*c*). These binding contacts are structurally interdependent so that a receptor-triggered structural change can initiate a cooperative unravelling of the interactions that secure the nucleotide. Following release of GDP, GTP binds to the open nucleotide-binding cleft. As already described, the newly bound γ-phosphate initiates a conformational change in the switch II region. This potentially alters the structure of the abutting C-terminal 5–10 residues, ultimately leading to dissociation from photoactivated rhodopsin and the $\beta\gamma$ heterodimer.

Effector activation

$G_{t\alpha}\cdot$GTP binds and activates cGMP–PDE by displacing the enzyme's inhibitory γ-subunits[7]. Two separate studies implicate the G_α surface loops ($\alpha2/\beta4$, $\alpha3/\beta5$ and $\alpha4/\beta6$) in effector binding and activation. All of these loops form a contiguous patch of exposed residues (Fig. 4*a*). Peptides designed to mimic the $\alpha3/\beta5$ loop and the $\alpha4/\beta6$ loop associate strongly with PDE-γ. In addition, the $\alpha4/\beta6$ loop peptide activates cGMP–PDE[30]. Work with Gsα/Giα chimaeras also establishes a critical role for all three loops in the control of adenylyl cyclase activity[31]. The crystal structure of $G_{t\alpha}\cdot$GTPγS suggests how GTP might induce a structural change in these three effector-binding loops. The crystal structure also provides a structural basis for a recently implicated fourth effector binding region of $G_{t\alpha}$ (J. Mills *et al.*, manuscript in preparation).

Structural consequences of GTP binding. A change in proteolysis sensitivity indicates that the switch II region (residues 198–215) undergoes a distinct transition, probably from a flexible structure to a well ordered helix upon binding GTP. Two $\alpha2$ residues, Arg 204 and Trp 207, are protected from tryptic[25] and chymotryptic[26] hydrolysis, respectively, in the $G_{t\alpha}\cdot$GTPγS complex but not in the $G_{t\alpha}\cdot$GDP complex. Moreover, GTP/GDP exchange alters the fluorescence of Trp 207. Trp 207 of the switch II region is coupled to the γ-phosphate through the peptide between two glycine residues, both of which are conserved in the heterotrimeric G-protein family; the backbone amide nitrogen of the mutationally sensitive Gly 199 (ref. 32) contacts the γ-phosphate while the carbonyl of that peptide (Gly 198) accepts a hydrogen bond from the heterocyclic —NH— of Trp 207. Thus, the first two turns of $\alpha2$ are stabilized by interaction with the γ-phosphate. This tandem of conserved glycines may provide the necessary freedom for a potential conformational change in $\alpha2$ upon loss of these interactions with the γ-phosphate. In addition, we predict that in the GDP-bound state the negative side chain of Glu 203 (like Asp 86 of EF-Tu–GDP[13–15]) will occupy the positively charged pocket vacated by the γ-phosphate and will be expelled upon GTP binding.

The GDP/GTP-induced changes in $\alpha2$ provide a direct link to changes in the $\alpha2/\beta4$ loop. These changes can, in turn, be transmitted to $\alpha3$ and $\alpha4$ through a series of ionic and van der Waals contacts (Fig. 4*b*). These interhelical contacts provide a contiguous link between the direct effect of the GTP's γ-phosphate on the $\alpha2$ helix and the effector-binding $\alpha3/\beta5$ and $\alpha4/\beta6$ loops. Many of the structurally observed linkages pictured in Fig. 4*b* involve residues that are also conserved in effector-targeted heterotrimeric G proteins and are not found in Ras (Fig. 2*b*), which reflects the importance of these stereochemical relays in the positioning of the three effector modulating loops.

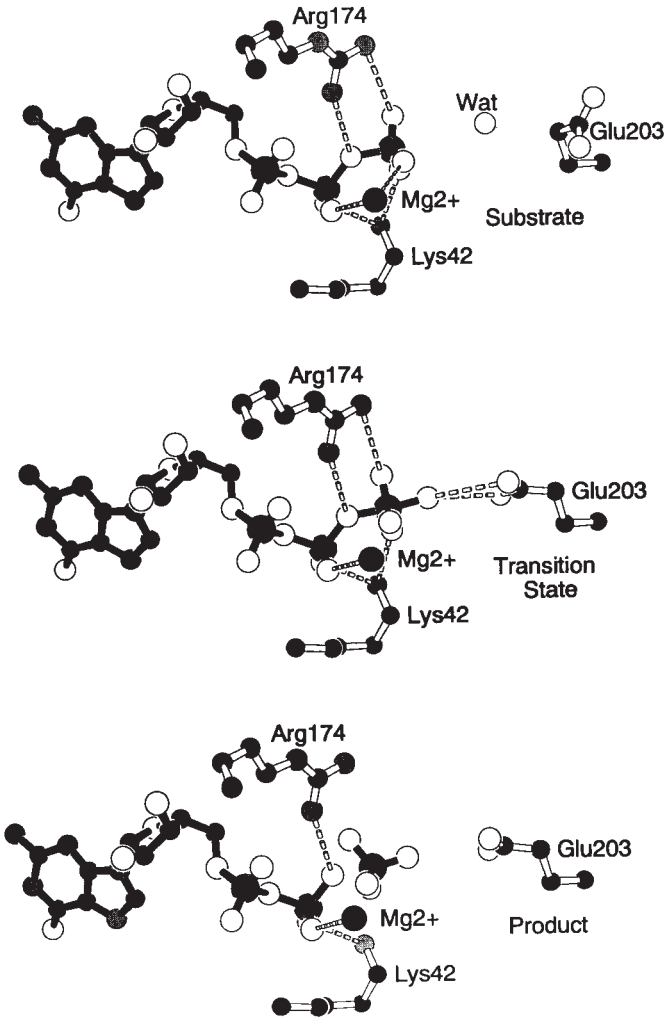


FIG. 5 Proposed catalytic mechanism for $\beta\gamma$ phosphate hydrolysis. The ground state (substrate), transition state, and products have been modelled to reflect a bound GTP rather than the catalytically resistant GTPγS. The only difference with the observed structure for the substrate is the P–S bond which has been shortened from 1.86 Å to 1.45 Å to reflect a typical P–O bond length. The active site water molecule in the ground state is situated directly in-line wth the γ-phosphate, but positioned 3.9 Å away from the phosphorus. This distance is the result of packing constraints imposed on the system by the bulky sulphur atom of the bound inhibitor (van der Waals radii is 1.7 Å versus 1.35 Å for oxygen). In the transition state, the oxygen atom of the crystallographically observed active site water has been moved to within bonding distance of the γ phosphorus. Rotation of Glu 203's Cβ–Cγ bond places its carboxylate in an ideal position to activate the water molecule by proton abstraction (general base catalysis). In addition, the interactions of the pentacoordinated phosphate with Lys 42, Arg 174 and Mg^{2+} serve to neutralize the negative charge developed in the transition state leading to enhanced catalysis through transition state stabilization. In the last step of the reaction (product), the bond between the γ-phosphorus and the bridging oxygen breaks leading to the release of a free phosphate and GDP. The likelyhood of a glutamate-activated water molecule serving as the attacking nucleophile in nucleotide hydrolysis has been suggested by Steitz[52].

The I2/α2/phosphate-binding loop linkage. One of the most striking stereochemical features of the $G_{t\alpha}\cdot GTP\gamma S$ structure are the salt bridges that link I2 (residues 226–239) to both the phosphate-binding region (residues 36–43) and α2 (Fig. 4*c*). These linkages probably make the position of an acidic cluster of side chains in the I2 segment (Asp 233, Asp 234 and Glu 235) sensitive to the presence or absence of the nucleotide's γ-phosphate. We speculate that this acidic patch as well as residues Glu 306, Asp 311 and Glu 314 (ref. 33) form an interface involved in binding the highly basic region of the γ-subunit of cGMP–PDE. A peptide corresponding to residues 232–259 binds PDE-γ (J. Mills *et al.*, manuscript in preparation). The sequence variability at positions corresponding to Asp 234 and Glu 235 in the heterotrimeric family of α-subunits suggests that this surface can provide specificity in effector interactions. On the other hand, the conservation of the ionic linkages to the switch II region reflects an important stereochemical connection between GTP binding and potential conformational changes in the I2 region (Fig. 2*b*).

GTP hydrolysis

There are two features in this crystal structure relevant to GTP hydrolysis that are not found in the structures of the stable GTP-analogue complexes of Ras[12] and EF-Tu[13]. The first is conserved arginine residue (Arg 174) that contacts the γ-phosphate directly. The second is a candidate for the general base (Glu 203) that triggers the hydrolytic attack. Arg 174 provides hydrogen bonds to the α- and γ-phosphates and the oxygen bridging the β- and γ-phosphates. These direct contacts are positioned to stabilize the negative charge that develops in the presumed pentacovalent transition state and thereby facilitate the release of the γ-phosphate from the oxygen bridging the β- and γ-phosphates (Fig. 5). Mutations of the corresponding arginine in $G_{i\alpha}$ (ref. 34) and $G_{s\alpha}$ (ref. 35) dramatically reduce their respective GTPase activities and constitutively activate their respective target effectors. Also, the guanidino group of this arginine is the target for ADP-ribosylation by cholera toxin, which blocks hydrolysis[18].

A water molecule is identically positioned to serve as a nucleophile for a direct in-line attack on the γ-phosphate in all three crystallographic representations of the $G_{t\alpha}\cdot GTP\gamma S$ complex (Fig. 5). This fixed water site is located in the corresponding position in the high-resolution structure of Ras·GMP.PNP[12] and EF-Tu·GMP.PNP[13]. It is likely that this water molecule has a nucleophilic role in the GTPase mechanism, but the structure of Ras has not identified the general base that activates it. In the recently described structure of EF-Tu[13], a histidine corresponding to Gln 61 in Ras or Gln 200 in $G_{t\alpha}$, is proposed to move after EF-Tu binds the ribosome in such a way that it is positioned to act as a general base. Although the mutational sensitivity of a corresponding glutamine in Ras and in G_α subunits implicates it in the GTPase mechanism, this structure and that of Ras offer little in the way of a structural explanation. In $G_{t\alpha}$, the side-chain carboxylate of Glu 203 is situated so that a rotation of its Cβ–Cγ bond places the negatively charged carboxylate in a position to abstract a proton and thereby activate the water (Fig. 5). This glutamate is well conserved across the large family of G_α subunits, which, coupled with our structural information, underscores its functional importance in heterotrimeric G-protein α-subunits (Fig. 2*b*). Whereas a corresponding acidic residue is absent in the smaller GTPases such as Ras, Glu 63 in the disordered α2 region of Ras (Asp 86 in EF-Tu), corresponding to $G_{t\alpha}$ residue 202, could serve as an alternative general base. In fact, Glu 63 is a site for a GTPase-inhibiting mutation in $p21^{Ha\text{-}ras}$ (ref. 36).

Glutamate 203 resides on the α2 switch helix of $G_{t\alpha}$ and is homologous to α2 of Ras[17] and EF-Tu[13], whose conformations are also sensitive to the presence or absence of the γ-phosphate. It is interesting to note that most heterotrimeric α-subunits possess a significantly higher basal GTPase rate than the small GTPases such as Ras[37]. We would argue that the higher GTPase rates in heterotrimeric G-protein α-subunits reflects (1) an inherently better-structured α2 helix, which provides a properly positioned general base without the aid of a GTPase-activating protein, and (2) the presence of a transition-state-stabilizing interaction with the side chain of a conserved arginine corresponding to Arg 174. As α2 is structurally coupled to effector binding sites, Glu 203 could be productively oriented by bound effectors[5,6].

Finally, we offer an explanation for the resistance to hydrolysis conferred by the γ-thiophosphorothioate. The bulky sulphur atom of GTPγS, identified unequivocally by its appearance in (*Fo*–*Fc*) maps and its standard phosphorus–sulphur bond distance (1.86 Å), sterically shields the phosphorus atom from a close approach by the active-site water molecule. Indeed, this prevents the side chain of Glu 203 from assuming the position needed to support in-line attack by the activated water molecule. In addition, it appears from modelling of the pentacovalent transition state that Arg 174 prevents the thiophosphate from reaching the transition state owing to a steric clash between the firmly anchored guanidino group and the equatorial sulphur atom.

In summary, this study of $G_{t\alpha}\cdot GTP\gamma S$ is a first step in developing a stereochemical context for a dynamic understanding of G-protein-coupled signal transduction through seven-helix transmembrane receptors and their cognate heterotrimeric G proteins. Until now the structural basis for experiments designed to extract structure/function correlates has been limited to that provided by other GTPases such as Ras and EF-Tu. Whereas $G_{t\alpha}$ shares many nucleotide-binding elements with the larger family of GTPases, its value as a model for G-coupled signal transduction lies in its unique structure/function relationships which are, at best, tenuously extrapolated from the models of Ras and EF-Tu. When viewed in the context of genetic, biochemical and dynamic studies on transducin and similar G-coupled systems, the present crystal structure serves as a necessary but only partial link between the functions of these G proteins and the underlying molecular mechanism by which they operate. Structural work on other complexes will complete and clarify the stereochemical picture. □

Received 3 November; accepted 23 November 1993.

1. Conklin, B. R. & Bourne, H. R. *Cell* **73,** 631–641 (1993).
2. Blumer, K. J. & Thorner, J. *A. Rev. Physiol.* **53,** 37–57 (1991).
3. Tang, W. J., Krupinski, J. & Gilman, A. G. *J. biol. Chem.* **266,** 8595–8603 (1991).
4. Berstein, G. *et al. Cell* **70,** 411–418 (1992).
5. Arshavsky, V. Y. & Bownds, M. D. *Nature* **357,** 416–418 (1992).
6. Stryer, L. & Bourne, H. R. *A. Rev. Cell Biol.* **2,** 391–419 (1986).
7. Pfister, C. *et al. Cell Sig.* **5,** 235–241 (1993).
8. Liebman, P. A., Parker, K. R. & Dratz, E. A. *A. Rev. Physiol.* **49,** 765–791 (1987).
9. Mazzoni, M. R., Malinski, J. A. & Hamm, H. E. *J. biol. Chem.* **266,** 14072–14081 (1991).
10. Jones, T. A. *J. appl. Crystallogr.* **11,** 268–276 (1978).
11. Brunger, A. T. *XPLOR Version* 3.1 *Manual* Yale Univ. (1993).
12. Pai, E. F. *et al. EMBO J.* **9,** 2351–2359 (1990).
13. Berchtold, H. *Nature* **365,** 126–132 (1993).
14. Jurnak, F. *Science* **230,** 32–36 (1985).
15. La Cour, T. F. M., Nyborg, J., Thirup, S. & Clark, B. F. C. *EMBO J.* **4,** 191–212 (1985).
16. Saraste, M., Sibbald, P. R. & Wittinghofer, A. *Trends biochem. Sci.* **15,** 430–434 (1990).
17. Milburn, M. V. *et al. Science* **247,** 939–945 (1990).
18. Van Dop, C., Tsubokawa, M., Bourne, H. R. & Ramachandran, J. *J. biol. Chem.* **259,** 696–698 (1984).
19. Hamm, H. E. *et al. Science* **241,** 832–835 (1988).
20. Dratz, E. A. *et al. Nature* **363,** 276–281 (1993).
21. Conklin, B. R., Farfel, Z., Lustig, K. D., Julius, D. & Bourne, H. R. *Nature* **363,** 274–276 (1993).
22. Simonds, W. F., Goldsmith, P. K., Codina, J., Unson, C. G. & Spiegel, A. M. *Proc. natn. Acad. Sci. U.S.A.* **86,** 7809–7813 (1989).
23. Gilman, A. G. *A. Rev. Biochem.* **56,** 615–649 (1987).
24. Van Dop, C. *et al. J. biol. Chem.* **259,** 23–26 (1984).
25. Fung, B. K.-K. & Nash, C. R. *J. biol. Chem.* **258,** 10503–10510 (1983).
26. Mazzoni, M. R. & Hamm, H. E. *J. Prot. Chem.* **12,** 215–221 (1993).
27. Higashijima, T. *et al. J. biol. Chem.* **262,** 752–756 (1987).
28. Kaziro, Y. *Biochem. biophys. Acta* **505,** 95–127 (1978).
29. Downward, J., Riehl, R., Wu, L. & Weinberg, R. A. *Proc. natn. Acad. Sci. U.S.A.* **87,** 5998–6002 (1990).
30. Rarick, H. M., Artemyev, N. O. & Hamm, H. E. *Science* **256,** 1031–1033 (1992).
31. Berlot, C. H. & Bourne, H. R. *Cell* **68,** 911–922 (1992).
32. Miller, R. T., Masters, S. B., Sullivan, K. A., Beiderman, B. & Bourne, H. R. *Nature* **334,** 712–715 (1988).
33. Artemyev, N. O. *et al. J. biol. Chem.* **268,** 23611–23615 (1993).
34. Gupta, S. K. *et al. J. molec. cell. Biol.* **12,** 190–197 (1992).
35. Landis, C. A. *et al. Nature* **340,** 692–696 (1989).
36. Gibbs, J. B., Sigal, I. S. & Scolnick, E. M. *Trends biochem. Sci.* **10,** 350–353 (1985).

37. Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **348,** 125–132 (1990).
38. Kabsch, W. *J. appl. Crystallogr.* **21,** 67–71 (1988).
39. Kabsch, W. *J. appl. Crystallogr.* **21,** 916–924 (1988).
40. Otwinowski, Z. *ML-PHARE CCP4 Proc.* 80–88 (Daresbury Laboratory, Warrington, UK, 1991).
41. Zhang, K. Y. *Acta crystallogr.* D**49,** 213–222 (1993).
42. Brunger, A. T. *Nature* **355,** 472–475 (1992).
43. Carson, M. *J. appl. Crystallogr.* **24,** 958–961 (1991).
44. Robishaw, J. D., Russell, D. W., Harris, B. A., Smigel, M. D. & Gilman, A. G. *Proc. natn. Acad. Sci. U.S.A.* **83,** 1251–1255 (1986).
45. Nukada, T. *et al. FEBS Lett.* **197,** 305–308 (1986).
46. Van Meurs, K. P. *et al. Proc. natn. Acad. Sci. U.S.A.* **84,** 3107–3111 (1987).
47. Strathmann, M. & Simon, M. I. *Proc. natn. Acad. Sci. U.S.A.* **87,** 9113–9117 (1990).
48. Casey, P. J., Fong, H. K. W., Simon, M. I. & Gilman, A. G. *J. biol. Chem.* **265,** 2383–2390 (1990).
49. Yatsunami, K. & Khorana, H. G. *Proc. natn. Acad. Sci. U.S.A.* **82,** 4316–4320 (1985).
50. Capon, D. J., Chen, E. Y., Levinson, A. D., Seeburg, P. H. & Goeddel, D. *Nature* **302,** 33–37 (1983).
51. Laursen, R. A., L'Italien, J. J., Nagarkatti, S. & Miller, D. L. *J. biol. Chem.* **256,** 8102–8109 (1981).
52. Steitz, T. A. & Story, R. M. *Proc. The Robert A. Welch Foundation Conference on Chemical Research* XXXVI Regulation of Proteins by Ligands 173–186 (Houston, 1992).

ACKNOWLEDGEMENTS. We thank Z. Otwinowski for his solvent-flattening routine using maximum entropy algorithms and for technical advice during data processing, G. Van Duyne for technical assistance during data collection, Y.-C. Kim for advice during the heavy-atom search, A. Garcia for preparing bovine rod outer segments and J. Mills for preliminary work on lanthinide substitutions in transducin. This work was supported by grants from the NIH to P.B.S. and H.E.H. H.E.H. was also supported by the American Heart Association and Research to Prevent Blindness. J.P.N. was an NSF Chemistry and NIH postdoctoral fellow. The coordinates of the $G_{t\alpha}\cdot GTP\gamma S$ complex will be submitted to the Brookhaven Protein Data Bank.

LETTERS TO NATURE

Pulsar motion effect and Geminga's high braking index

G. S. Bisnovatyi-Kogan* & K. A. Postnov†

* Space Research Institute, Profsoyuznaya st. 84/32, 117810 Moscow, Russia
† Sternberg Astronomical Institute, Universitetskij pr. 13, 119899 Moscow, Russia

THE pulsar braking index, *n*, is a dimensionless quantity describing the rate at which a magnetized neutron star loses rotational energy[1]. It can be determined from pulsar timing measurements, and for distant pulsars is found to lie close to $n=3$ (ref. 2), as predicted by theoretical models of pulsar emission mechanisms[3–5]. In contrast, the timing parameters—in particular the second derivative of the pulsation frequency—of the nearby pulsar Geminga[6–11] indicate an extremely large braking index of about 10–30. To understand this property of Geminga, we consider here the effect on the measured timing parameters of a pulsar's motion through space. We find that the Doppler effect alone can give a high apparent braking index, but only if the pulsar is very close and has an abnormally high velocity (>1,000 km s^{-1}). A more likely (but related) cause of the high braking index is the pulsar's proper motion: failure to correct for changes in the source coordinates with time can greatly influence the higher derivatives of the pulsar frequency, and lead to an erroneous value of *n*. A self-consistent analysis of the timing data, which account for both the proper motion and the Doppler effect, should permit a reliable estimate of the distance to Geminga.

The value of the pulsar braking index, n, characterizes the dependence of the energy loss of a neutron star on its angular velocity, Ω (or frequency $f=\Omega/2\pi$). Energy losses due to magnetic dipole radiation[3,4] or to stellar wind caused by the induced electrical field of the rotating magnetic dipole[5] lead to the dependence $\dot{E}\propto B^2\Omega^4$ (where E is the rotational energy and B the surface magnetic field of the neutron star). Given that $E\propto\Omega^2$, one gets $\dot{\Omega}\propto\Omega^3$, which leads to $\ddot{\Omega}/\dot{\Omega}=3\dot{\Omega}/\Omega$ and the canonical value of the braking index $\ddot{\Omega}\Omega/\dot{\Omega}^2=3$. The braking indices for most well studied pulsars are <3 (ref. 2). These pulsars are rather distant and their space motion does not affect significantly their braking indices. Recent X-ray[6] and γ-ray[7–10] observations of the pulsar Geminga (2CG195+04) and an analysis of earlier data from the COS-B satellite[11] gave the following values for frequency f_0 and its first (f_1) and second (f_2) derivatives: $f_0\approx4.2178$ Hz, $f_1\approx-1.952\times10^{-13}$ Hz s^{-1} $f_2=(2.8\pm1.6)\times10^{-25}$ Hz s^{-2} (1σ deviation), implying that $n\approx10$–30. To reduce this to the canonical value of the braking index, the value of f_2 would need to be an order of magnitude smaller, that is, $f_2^*=3f_1^2/f_0\approx2.7\times10^{-26}$ Hz s^{-2}.

A pulsar's motion in space could potentially influence the observed timing parameters f_0, f_1 and f_2 in two ways. The first effect is the well known Doppler effect. The second effect arises when the pulsar's source coordinates (used for making the barycentric correction) are incorrect, for example, when they have been measured only once and no information about the proper motion of the source has subsequently been obtained. The second derivative of the frequency, f_2, is particularly sensitive to the accuracy of the source coordinates.

Consider the influence of the space motion on the braking index in more detail, beginning with the Doppler effect. Let a pulsar with proper frequency f move with respect to the Earth with a velocity v at a distance l. Then the observed pulsar frequency f is

$$f=\hat{f}\left(1+\frac{v}{c}\cos\theta\right) \tag{1}$$

where c is the speed of light and θ is the angle between the pulsar velocity and the direction of propagation of the radiation in the observer's frame. Expanding $\hat{f}(t)$ and $f(t)$ to the second order $\hat{f}=\hat{f}_0+\hat{f}_1t+\frac{1}{2}\hat{f}_2t^2$ and differentiating with respect to time t, one gets

$$f_0=\hat{f}_0\left(1+\frac{v}{c}\cos\theta\right) \tag{2}$$

$$f_1=\hat{f}_1\left(1+\frac{v}{c}\cos\theta\right)-\hat{f}_0\frac{v^2}{cl}\sin^2\theta \tag{3}$$

$$f_2=\hat{f}_2\left(1+\frac{v}{c}\cos\theta\right)-2\hat{f}_1\frac{v^2}{cl}\sin^2\theta - 3\hat{f}_0\frac{v^3}{cl^2}\sin^2\theta\cos\theta \tag{4}$$

For realistic values of v, the term in brackets has little effect on the values of frequency and its derivatives, and we will therefore put it equal to unity. The apparent first derivative f_1 always becomes more negative due to pulsar motion, even if the pulsar has constant proper frequency, as was first noted by Shklovskij[12]. In contrast, the second derivative f_2 will be increased by the Doppler effect if the pulsar moves away from the observer ($\cos\theta<0$). For typical values of $v/c\approx10^{-3}$ (that is, for pulsar velocities of $\sim$300 km s^{-1}) and adopting the measured Geminga values for $\hat{f}_1$ and $\hat{f}_2$, the second term in equation (4) is much less than the third; but both make the apparent value of f_2 higher.

The observer measures an apparent braking index

$$n=\hat{n}\frac{1+\frac{2}{\hat{n}}\left(\frac{2\tau_d v}{l}\right)\left(\frac{v}{c}\right)\sin^2\theta-\frac{3}{\hat{n}}\left(\frac{2\tau_d v}{l}\right)^2\left(\frac{v}{c}\right)\sin^2\theta\cos\theta}{\left[1+\left(\frac{2\tau_d v}{l}\right)\left(\frac{v}{c}\right)\sin^2\theta\right]^2} \tag{5}$$

where $\tau_d\equiv-f_0/2f_1$ is the pulsar dynamical age and $\hat{n}$ is the true braking index defined in terms of $\hat{f}_i$. If the proper values of $\hat{f}_1$ and $\hat{f}_2$ are so small that only the second term in equation (3) and the third term in equation (4) are significant, the apparent

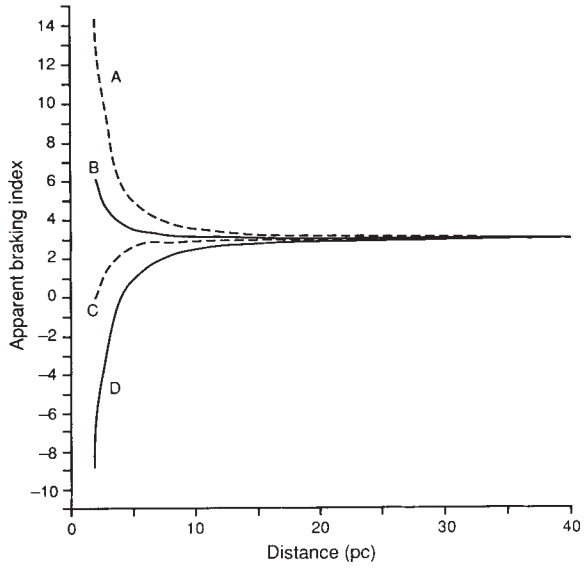


FIG. 1 Apparent braking index n of Geminga as a function of distance to the Earth l for pulsar space velocity 300 km s^{-1} and different angles θ; trace A, $\cos\theta = -0.6$; trace B, $\cos\theta = -0.95$; trace C, $\cos\theta = +0.95$; trace D, $\cos\theta = +0.6$. (The pulsars dynamical age τ_a is taken as 3.6×10^5 yr.)

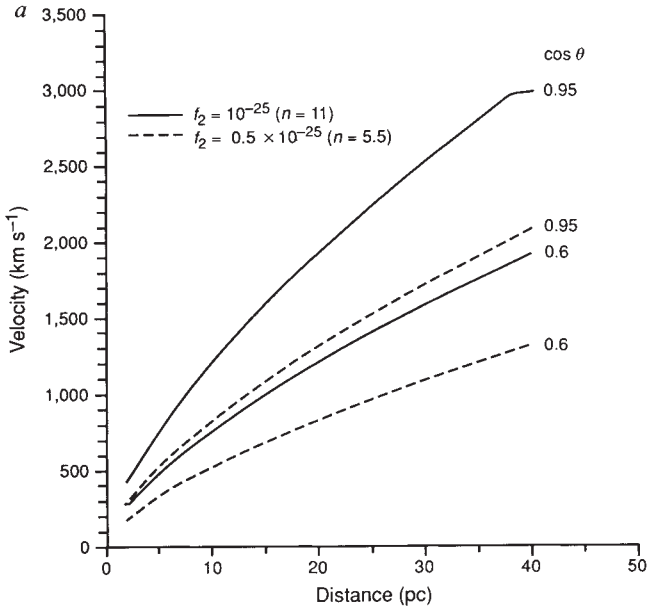
braking index can be very high, $n \approx -3(c/v)\cos\theta \sin^{-2}\theta$, but this is not the case for Geminga. Figure 1 shows how n changes with distance to pulsar with $\tau_d = 3.6 \times 10^5$ yr (the dynamical age of Geminga) and $v/c = 10^{-3}$.

By adopting the canonical value for the true braking index, and using the observed f_i, equation (5) relates the pulsar velocity v to the distance l. The pulsar velocities which give $n = 11$ and $n = 5.5$ (corresponding to the values of f_2 at the lower 1σ and 1.5σ limits respectively) are plotted in Fig. 2*a* as function of the distance to Geminga for $\cos\theta = -0.6$, at which the Doppler effect is maximal, and for $\cos\theta = -0.95$. As can be seen in Fig. 2*a*, the value of 40 pc suggested[7] as the distance to Geminga requires a very high pulsar velocity, $v > 1{,}000$ km s^{-1}. Spatial velocities this high cannot be obtained in standard models of pulsar formation, in which it is assumed that the moderately high velocities of some young pulsars result from disruption of progenitor binary systems during the supernova explosion, so that a 'kick' velocity during collapse should be involved. Some recently proposed mechanisms for collapse anisotropy[13,14] may also impart a velocity to the pulsar, but it is not clear whether this will be sufficient. Nevertheless, it should be noted that at least two pulsars with velocities $v > 1{,}000$ km s^{-1} have now been discovered: PSR1757−24 (ref. 15) and PSR2224+65 (refs 16, 17).

A straightforward observational test for a high pulsar velocity can be made by measuring the proper pulsar motion, $\mu = 2''\ yr^{-1} \times (10\ pc/l)(v/100\ km\ s^{-1}) \sin\theta$. Proper motions of Geminga corresponding to the velocities from Fig. 2*a* are presented in Fig. 2*b*.

Limits on the observed proper motion can be obtained from the precise X-ray identification of Geminga. The best accuracy to date is the 3.2″ error circle given by observations using the High Resolution Imager (HRI) on the Einstein satellite in 1981, which has been used for Solar System barycentric correction of the pulse arrival times. The 25 mag to star *G*″, which has been suggested as the optical counterpart of Geminga[18,19], and also used[9] for the barycentric correction, lies within the Einstein error box[18,19]. During Rosat observations[6] in 1991, photons have been gathered from inside a 1.5′ error circle. Combined with the Einstein observations, this implies an upper limit of ~0.15′ yr^{-1} for Geminga's proper motion.

Finding proper motion of faint stars inside the X-ray error box described above could be strong evidence for an association with Geminga. A search for the proper motion of the star *G*″ has recently been done by Bignami *et al.*[20] in the optical band. Evidence for proper motion of 0.2″ yr^{-1} has been obtained, but this is much smaller than our estimate from consideration of the Doppler effect alone. Note, that in the deep optical exposure of the area around Geminga in 1984 (ref. 18), the optical counterpart could have been missed if the proper motion exceeds 1″ yr^{-1}. This will bring it out of the X-ray error box in 3 years. The direct measurement of the X-ray position with an accuracy of several arcseconds is desirable. Independent search for proper motion of Geminga could be made in the X-ray band by Rosat High Resolution Imager telescope. While the optical identification of Geminga still remains uncertain, we now show that even



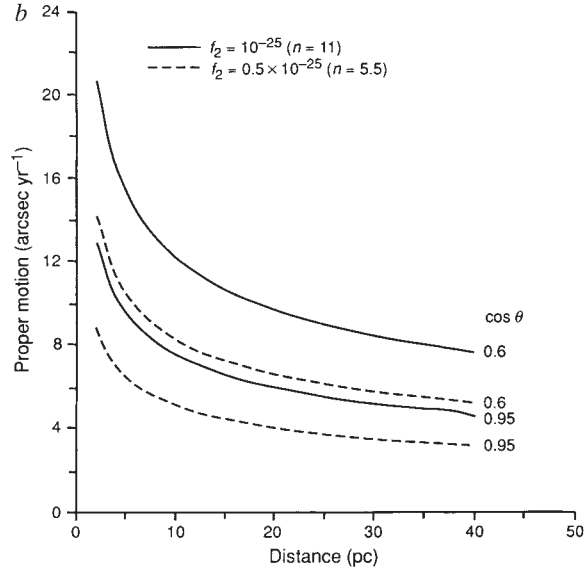


FIG. 2 *a*, Geminga space velocity v as a function of distance to Earth l for different angles θ and values of the apparent second derivative of the pulsar frequency f_2, assuming a canonical value of 3 for the Geminga braking index. *b*, Expected Geminga proper motion μ for the velocities from *a*.

the small paper motion of G'' could be sufficient for obtaining smaller f_2 and a realistic braking index, after removing the errors it introduces into the barycentric correction.

The errors in Geminga coordinates caused by the pulsar's proper motion influence the values obtained from timing analysis for f_0, f_1 and especially f_2 and n. The use of slightly different coordinates for barycentric correction (the positions of the Einstein HRI error circle and G'') has shown[10] large sensitivity of f_2 to the coordinates used. An estimate of this effect can be made by a simplified procedure of barycentric correction which only accounts for the motion of the Earth around the Sun. This yields an approximate formula for the error $\Delta\ddot{P}_e$ in measuring the second derivative of the pulsar period P due to the pulsar location error $\Delta\lambda$ of the form

$$\Delta\ddot{P}_e \approx -\frac{PR\omega^3}{c}\Delta\lambda\cos(\lambda-\omega t) \qquad (6)$$

Here $R=1$ AU, $\omega=2\pi/1\ \mathrm{yr} \ll 1/P$. (This formula was obtained in collaboration with S. Repin. A detailed investigation of the different effects connected with barycentric corrections will be published elsewhere.) The values of the errors $\Delta\ddot{P}_e$ as well as ΔP_e and $\Delta\dot{P}_e$ vary with time, and thus the accuracy of the values of P, $\dot{P}$ and $\ddot{P}$ obtained by averaging will depend on the duration of the observations. We can, however, use equation (6) to define an upper limit on the errors imposed on the second derivative, giving $\Delta f_{2e} \approx 8\times10^{-23}(\Delta\lambda)$ where $\Delta\lambda$ is expressed in arcsec. The abnormally large value of f_2 obtained in ref. 11 is much smaller than this upper limit on Δf_{2e} for a coordinate error of $\Delta\lambda=$ 1 arcsec, and in principle this could be explained by the proper motion of G''. This value of Δf_{2e} is also comparable to the upper limit for f_2 obtained[10] using the Egret instrument on the Compton Gamma Ray Observatory (GRO), where a much shorter observational time was used.

As mentioned above, the real correction Δf_{2e} depends on the observational interval T. For continuum observations and constant error $\Delta\lambda$, Δf_{2e} drops exponentially when $T \gg 1$ yr. But when the proper motion is significant, $\Delta\lambda=\Delta\lambda(t)$, and even long-term observations will not smear out the error Δf_{2e}. It is also clear that the barycentric correction required by the proper motion overwhelms the Doppler effect (third term in the equation (4)) even for errors in the pulsar location as small as 1 arcsec. The maximum possible influence by the motion of the GRO satellite around the Earth in presence of the location error $\Delta\lambda$ could be even higher, $\Delta f_{2s} \approx (v_a/c)(\omega_a^2/P)\Delta\lambda \gg \Delta f_{2e}$ (here ω_a is the angular frequency of the apparatus rotation around the Earth, $\omega_a \gg \omega$, and v_a is the satellite speed on the orbit). However, the exponential decay in the error due to time averaging is a much stronger effect here, so the errors introduced by the satellite's orbit are probably averaged out. Nevertheless, the numerical analysis must be done for checking these semiquantitative estimations.

We conclude that a self-consistent procedure of barycentric correction which accounts for both the pulsar proper motion and the Doppler effect will give reasonable values of n, f_2, v, and should therefore permit a reliable estimation of the distance to Geminga, amd improved accuracy of its angular coordinates. □

Received 11 May; accepted 22 October 1993.

1. Manchester, R. N. & Taylor, J. H. *Pulsars* (Freeman, San Francisco, 1977).
2. Manchester, R. N., Durdin, J. M. & Newton, L. M. *Nature* **313,** 374–376 (1985).
3. Pacini, F. *Nature* **216,** 567–568 (1967).
4. Gunn, H. E. & Ostriker, J. P. *Nature* **221,** 454–455 (1969).
5. Goldreich, P. & Julian, W. H. *Astrophys. J.* **157,** 869–875 (1969).
6. Halpern, J. P. & Holt, S. S. *Nature* **357,** 222–223 (1992).
7. Bertsch, D. L. *et al. Nature* **357,** 306–307 (1992).
8. Bignami, F. G. & Caraveo, P. A. *Nature* **357,** 287 (1992).
9. Mattox, J. R. *et al. IAU Circ. No.* 5583.
10. Mayer-Hasselwander, H. A. *et al. IAU Circ. No.* 5649.
11. Hermsen, W. *et al. IAU Circ. No.* 5541.
12. Shklovskij, I. S. *Astr. Zh.* **46,** 715–719 (1969).
13. Bisnovatyi-Kogan, G. S. & Moiseenko, S. G. *Astr. Zh.* **69,** 563–571 (1992).
14. Bisnovatyi-Kogan, G. S. *Astr. Astrophys. Trans.* **4,** 287–294 (1993).
15. Frail, D. A. & Kulkarni, S. R. *Nature* **352,**785–787 (1991).
16. Harrison, P. A., Lyne, A. G. & Anderson, B. in *Proc. NATO-ARW, X-ray Binaries and Recycled Pulsars* (eds Van den Heuvel, E. & Rappaport, S.) 155–160 (Kluwer, Dordrecht, 1992).
17. Cordes, J. M., Romani, R. W. & Lundgren, S. C. *Cornell Univ. preprint* Dec. 1992.
18. Bignami, G. F., Caraveo, P. A., Paul, J. A., Salotti, L. & Vigroux, L. *Astrophys. J.* **319,** 358–364 (1987).
19. Halpern, J. P. & Tytler, D. *Astrophys. J.* **330,** 201–220 (1988).
20. Bignami, G. F., Caraveo, P. A. & Mereghetti, S. *Nature* **361,** 704–707 (1993).

ACKNOWLEDGEMENTS. G.S.B.-K. is grateful to G. F. Bignami for attracting his attention to the unusually high values of n and f_2 of Geminga.

Role of sp^3 carbon and 7-membered rings in fullerene annealing and fragmentation

Robert L. Murry, Douglas L. Strout, Gregory K. Odom & Gustavo E. Scuseria*

Department of Chemistry and Rice Quantum Institute, Rice University, Houston, Texas 77251-1892, USA

WHEN fullerenes[1,2] are fragmented by laser irradiation, they lose C_2 fragments and retain a closed carbon cage[3]. The detailed mechanism of this process remains unknown, although survival of the cage implies that annealing (rearrangement of the bonding) must play an important role[3,4]. Here we use *ab initio* quantum-chemical calculations to show that fullerene annealing happens more readily than fragmentation, and that both are intimately related. Our findings imply that the assumptions commonly made about fullerenes[5]—that they are composed of five- and six-membered rings of sp^2 carbons— are not valid under high-energy conditions. In particular, the appearance of sp^3 carbon and seven-membered rings is central in both the annealing and fragmentation processes. Our theoretical predictions imply that the high-energy processes of fullerene growth[6–11] and coalescence[12] are much richer than previously thought, and that their mechanisms may also involve structures containing sp^3 carbon and seven-membered rings. Our results may aid in the design of experimental methods for controlling the nature of fullerene cages (for example, doping, opening and re-closing them).

Quantum-mechanical calculations in this work include the semi-empirical MNDO procedure[13], the *ab initio* direct SCF Hartree–Fock method[14,15], and local (LDA)[16] and non-local (BLYP)[17] density functional theory. The SCF method has been quite reliable in predicting fullerene structures[18,19], and the BLYP method has been shown recently to reproduce the experimental thermochemistry of chemical reactions to within a few kcal mol^{-1} (ref. 20). The gradient-corrected density functional calculations presented here were obtained with our recently developed program (G.K.O. and G.E.S., manuscript in preparation). The carbon basis sets employed are identical to those used in previous *ab initio* fullerene studies[21,22]. We find that all theoretical tools employed in this work agree qualitatively on the main aspects of our model, with only relatively minor quantitative differences.

A relatively low-energy annealing mechanism would be the simplest explanation for many of the phenomena involved in fullerene rearrangement and fragmentation[4]. (The structures of A–I in the following discussion are given in Fig. 1.) The Stone–Wales mechanism[23] (A–C) is a possible route to rearrange polygons in the fullerene surface. Our theoretical predictions indicate that this reaction can occur through a nearly-planar transition state (B) with a significant activation barrier, calculated as 8.0 eV at the highest level of theory ($4s2p1d$/BLYP) employed here. Our own search for alternative processes to the Stone–Wales rearrangement was unsuccessful. We did find, however, that this rearrangement occurs through a different route to the one originally suggested. This new process involves a non-planar intermediate (D) which has one sp^3, and one sp hybridized carbon.

* To whom correspondence should be addressed.

Unlike the in-plane process where two bonds are broken simultaneously, our process involves two sequential 1,2 carbon shifts (migration of a bond from one carbon atom to an adjacent one) and has a lower activation barrier because bonds are broken and re-formed one at a time. This new process for the Stone–Wales rotation was thoroughly investigated both in model systems and in actual fullerenes such as C_{60}. In all cases we find results which are qualitatively consistent; the activation barrier through the sp^3 intermediate is lower than the one obtained through the in-plane route. As presented in Table 1, in the simplest model system for which the two processes are operative (the minimum sheet as shown in Fig. 1, and in columns three and four in Table 1), our best theoretical prediction for the activation barrier of the Stone–Wales annealing process obtained at the $4s2p1d$/BLYP level of theory (6.1 eV) is 1.9 eV lower than the in-plane mechanism. When an additional rim of carbon atoms is added to this minimum sheet in a way that creates an extended section of the C_{60} fullerene surface (see fifth and sixth columns in Table 1), the activation barrier for annealing through the sp^3 intermediate remains substantially lower than the in-plane process for all levels of theory. It should be expected that different fullerenes should have different barriers from those reported in Table 1. We have consistently found a shallow region of the potential energy surface where an sp^3 structure exists (see results for C_{60} below), and that the sp^3 annealing channel is lower in energy than the in-plane process.

Most remarkably, the newly discovered sp^3 intermediate provides a natural link between the processes of fullerene annealing and fragmentation. As shown in the bottom half of Fig. 1 (structures E–I), the sp^3 intermediate (D) provides the fullerene ground state (A) with two fragmentation pathways. Our calculations show that loss of C_2 proceeds by a series of intermediates that we will refer to as 'handles' (F) and 'sticks' (E, G and H). These interesting structures represent deep minima in the potential energy surface, but lie at relatively high energies. The presence of dangling bonds is expected to make them very reactive. We have located transition states at the MNDO level of theory (characterized by analytical second derivatives) for all the paths depicted in Fig. 1, both for model systems and for C_{60}. An important consequence of the novel fragmentation mechanism presented in Fig. 1 is that the C_2 molecule that leaves the fullerene surface is actually a bond shared by a hexagon and a pentagon (5–6 bond). This reveals a new type of C_2 loss process which does not require the creation of a pentagon–pentagon (5–5) defect (as originally suggested[3]) before C_{60} or the higher fullerenes can lose C_2. Moreover, the 5–6 C_2-loss process results in the appearances of a seven-membered ring (I) in the fullerene surface.

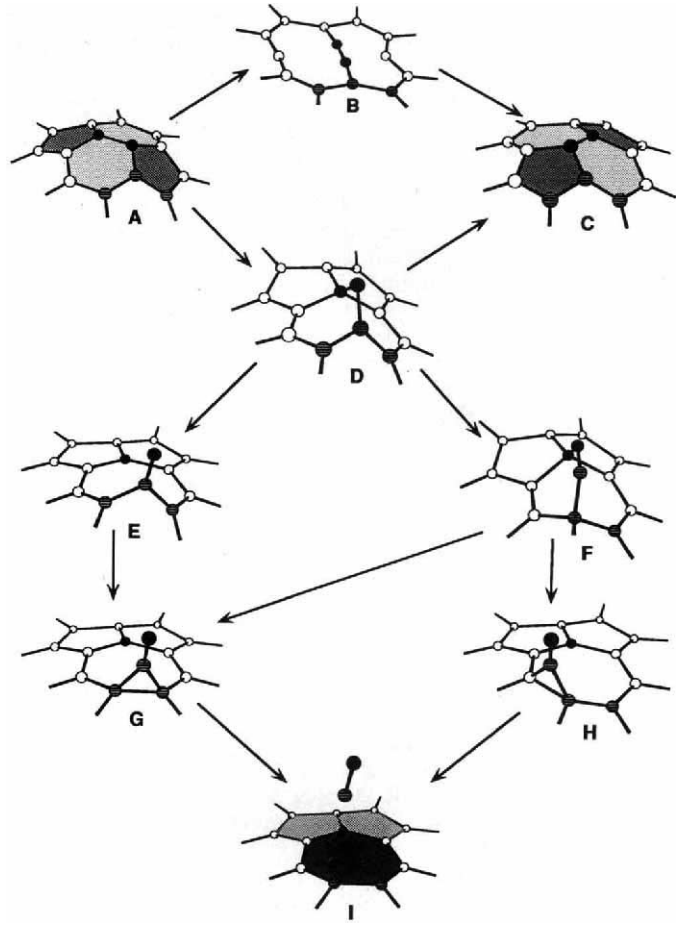
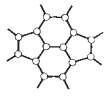
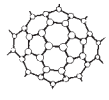


FIG. 1 The Stone–Wales rearrangement and its role in annealing and fragmentation. Atoms and polygons are shaded for ease in following these transformations. A, Starting fullerene section. B, In-plane Stone–Wales transition state. C, Ending section after a Stone–Wales rearrangement. If A is a portion of the C_{60} surface, then rearranging the pentagons and hexagons to create C will give a C_{60} structure with a pair of 5–5 defects. D, sp^3 intermediate. Note its role in connecting fullerene annealing (above) with fragmentation (below). E, Open C_2 stick. F, Handle. G, H, Closed C_2 sticks. I, End result after C_2 loss from a 5–6 bond, including seven-membered ring.

TABLE 1 Energetics of Stone–Wales annealing rearrangement for fullerene model systems

		Barriers (eV)			
					
		Minimum*		Extended†	
Method	Basis set	In-plane	sp^3 intermediate	In-plane	sp^3 intermediate
MNDO		8.6	5.9	8.2	5.9
SCF	$2s1p$	11.3	5.5	11.6	7.0
SCF	$4s2p$	9.6	7.8	9.6	6.3
LDA	$4s2p$	8.7	6.4	9.3	7.4
BLYP	$4s2p$	8.4	6.4	8.9	7.1
SCF	$4s2p1d$	9.5	7.1	9.5	6.1
LDA	$4s2p1d$	8.1	6.1	8.7	7.1
BLYP	$4s2p1d$	8.0	6.1	8.4	6.9

* Calculations on pyracylene-type model (see also Fig. 1). Individual structures optimized at the MNDO and SCF levels of theory for each basis set, respectively.

† Calculations on minimum model surrounded by a rim of additional hexagons and pentagons as found on the C_{60} surface. Individual structures optimized at the $2s1p$/SCF level of theory.

In Fig. 2 we present the results of a detailed investigation of the C_{60} potential energy surface. All geometries were optimized employing the MNDO method, and at these geometries energies were then obtained employing the $4s2p$ SCF, LDA and BLYP levels of theory. Transition states are indicated by 't.s.,' whereas all other structures correspond to true minima in the potential energy surface. The activation barriers for each process may be obtained directly from the relative energies listed below each structure in the figure. The bottom branch of Fig. 2 shows the processes involving the handle and stick isomers that lead to 5–6 C_2 loss and a C_{58} isomer containing a seven-membered ring (C_{58}(7-m.r.)). These fragmentation paths are analogous to those presented in Fig. 1. Remarkably, the presence of a seven-membered ring does not penalize the fullerene structure as much as one might believe, for C_{58}(7-m.r.) is only ~0.1 eV higher in energy than the predicted lowest-energy isomer of C_{58} (which has C_s symmetry (ref. 24) and is shown as $C_{58}(C_s)$ in Fig. 2). Furthermore, C_{58}(7-m.r.) may undergo a modified Stone–Wales rotation[25] and become the $C_{58}(C_s)$ isomer. Moreover, seven-membered rings have been observed recently in carbon nanotubes[26]. The top branch of Fig. 2 shows the two mechanisms

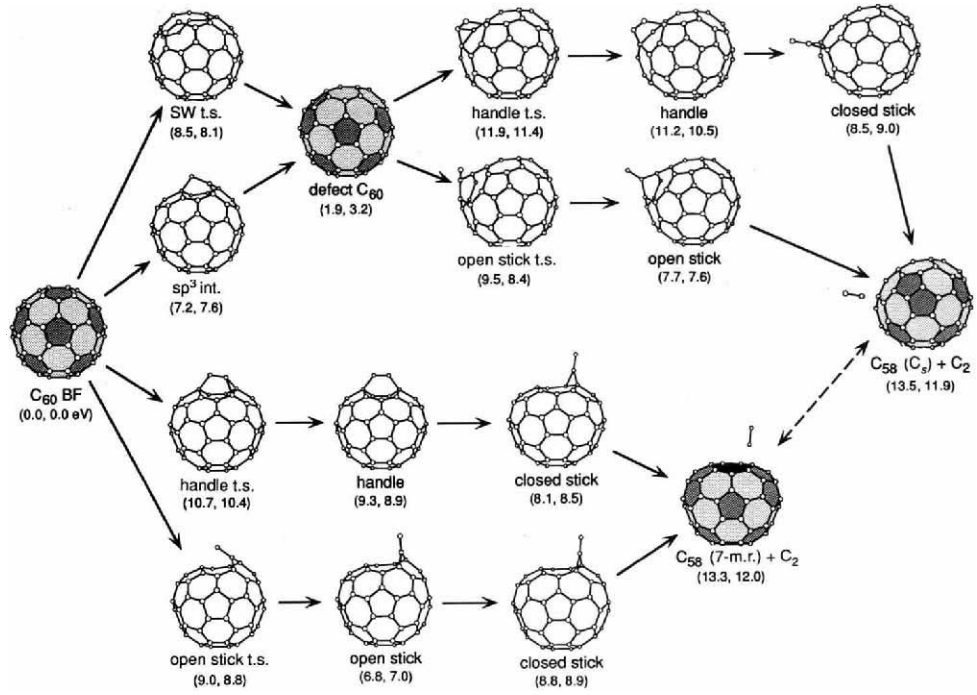


FIG. 2 Calculated annealing and fragmentation pathways from C_{60} buckminsterfullerene (BF). Geometries were optimized with the MNDO method (using the Gaussian 92, Rev. C program (M. J. Frisch *et al.*, personal communication)) and at these geometries energy calculations at the $4s2p$ SCF, LDA and BLYP levels of theory were carried out. The energies (in eV) of each structure relative to C_{60} BF (calculated at the MNDO and $4s2p$/BLYP levels of theory respectively) are shown in parenthesis. This map is not exhaustive, and only a few selected transition states (t.s.) are shown. Note that the C_{58} end products of C_2 fragmentation are connected through a modified Stone–Wales rearrangement (broken line, transition state not shown). All processes may occur reversibly. Abbreviations: int., intermediate; SW, Stone-Wales; C_{58} (7-m.r.), C_{58} isomer containing a seven-membered ring.

of Stone–Wales rotation leading to C_{60} with two 5–5 defects. The defect C_{60} may progress towards its own set of handles and sticks, which can in turn lose C_2 (from a 5–5 bond) and go to $C_{58}(C_s)$.

Note that all annealing processes in Fig. 2 have much lower barriers than the fragmentation processes and that the sp^3 intermediate lies energetically below the in-plane Stone–Wales transition state. In the case of C_{60}, the barrier to annealing through the sp^3 intermediate is much larger than in the model systems (MNDO, 7.2 eV; $4s2p$/BLYP; 7.6 eV; Fig. 2 and Table 1). We ascribe this effect to the deep-minimum character of C_{60} buckminsterfullerene and to its resilience to fragmentation. But geometry reoptimization at the *ab initio* levels of theory may lower these barriers as it will affect more strongly the high-lying transition states than the C_{60} deep minimum. Most importantly, the qualitative features of the C_{60} potential energy surface remain the same at the MNDO, SCF, LDA and BLYP levels of theory. This agreement between semi-empirical and more robust *ab initio* methods is strong evidence that our proposed mechanisms have a firm basis.

The top and bottom fragmentation channels of Fig. 2 have similar dissociation energies (11.9 and 12.0 eV) at the $4s2p$/BLYP level. The largest activation barriers for the respective multi-step processes of C_2 fragmentation are 8.4 and 8.9 eV. Thus, in actual photofragmentation processes, the two channels should both be active and give C_{58}(7 m.r.) and $C_{58}(C_s)$, a result which might allow experimental confirmation by designing probes to test the C_{58} product distribution. The relatively high barriers for these mechanisms are consistent with a view where C_{60} and the fullerenes are remarkably resilient[3], but nevertheless do fragment and anneal under high-energy conditions[9–12].

The prospect of capturing experimentally some of these intermediates seems exciting, for they are minima on the potential energy surface and should be metastable once formed. Our proposals could be tested experimentally by putting radical-scavenging species such as NO, SO_2 or O_2 into a cell containing C_{60} clusters photoexcited just below the photofragmentation threshold energy. Our proposed mechanisms suggest that under such conditions, C_{60} should be chemically reactive at the dangling bond sites of the handles, sticks and sp^3 intermediate. Mass-spectrum and photofragmentation (after further laser irradiation) analysis of these reaction products should yield insights into the nature of the structures that C_{60} 'samples' while photoexcited at high energies. □

Received 26 July; accepted 26 October 1993.

1. Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. *Nature* **318,** 162–163 (1985).
2. Krätschmer, W., Lamb, L. D., Fostiropoulos, K. & Huffman, D. R. *Nature* **347,** 354–358 (1990).
3. O'Brien, S. C., Heath, J. R., Curl, R. F. & Smalley, R. E. *J. chem. Phys.* **88,** 220–230 (1988).
4. Curl, R. F. *Phil. Trans. R. Soc.* A**343,** 19–32 (1993).
5. Kroto, H. W. *Nature* **329,** 529–531 (1987).
6. Kroto, H. W. & McKay, K. *Nature* **331,** 328–331 (1988).
7. Smalley, R. E. *Acct. Chem. Res.* **25,** 98–105 (1992).
8. Heath, J. R. in *Fullerenes: Synthesis, Properties, and Chemistry of Large Carbon Clusters* (eds Hammond, G. S. & Kuck, V. J.) 1–27 (Am. chem. Soc. Symp. Ser. No. 481, Washington DC, 1991).
9. Helden, G. V., Gotts, N. G. & Bowers, M. T. *Nature* **363,** 60–63 (1993).
10. Hunter, J., Fye, J. & Jarrold, M. F. *Science* **260,** 784–786 (1993).
11. McElvany, S. W., Ross, M. M., Goroff, N. S. & Diederich, F. *Science* **259,** 1594–1596 (1993).
12. Yeretzian, C., Hansen, K., Diederich, F. & Whetten, R. L. *Nature* **359,** 44–47 (1992).
13. Dewar, M. S. & Thiel, W. *J. Am. chem. Soc.* **99,** 4907–4917 (1977).
14. Almlöf, J., Faegri, K. Jr & Korsell, K. *J. comput. Chem.* **3,** 385–399 (1982).
15. Ahlrichs, R., Bär, M., Häser, M., Horn, H. & Kölmel, C. *Chem. Phys. Lett.* **162,** 165–169 (1989).
16. Parr, R. G. & Yang, W. *Density-Functional Theory of Atoms and Molecules* (Oxford Univ. Press, New York, 1989).
17. Scuseria, G. E. *J. chem. Phys.* **97,** 7528–7530 (1992).
18. Scuseria, G. E. in *Buckminsterfullerenes* (eds Billups, W. E. & Ciufolini, M. A.) 103–124 (VCH, New York, 1993).
19. Guo, T. *et al. Science* **257,** 1661–1664 (1992).
20. Johnson, B. G., Gill, P. M. W. & Pople, J. A. *J. chem. Phys.* **98,** 5612–5626 (1993).
21. Scuseria, G. E. *Chem. Phys. Lett.* **176,** 423–427 (1991).
22. Scuseria, G. E. *Chem. Phys. Lett.* **180,** 451–455 (1991).
23. Stone, A. J. & Wales, D. J. *Chem. Phys. Lett.* **128,** 501–503 (1986).
24. Shang, B. L., Wang, C. Z., Ho, K. M., Xu, C. H., Chan, C. T. *J. chem. Phys.* **97,** 5007–5011 (1992).
25. Saito, R., Dresselhaus, G & Dresselhaus, M. S. *Chem. Phys. Lett.* **195,** 537–542 (1992).
26. Iijima, S., Ichihashi, & Ando, Y. *Nature* **356,** 776–778 (1992).

ACKNOWLEDGEMENTS. G.E.S. is a Camille and Henry Dreyfus teacher-scholar. This research was supported by the US National Science Foundation and the Welch Foundation.

The interatomic structure of water at supercritical temperatures

P. Postorino*, R. H. Tromp†, M-A. Ricci‡, A. K. Soper§ & G. W. Neilson†||

* Dipartimento di Fisica, Università degli Studi di Roma La Sapienza, Piazzale Aldo Moro 2, I-00185 Rome, Italy
† H.H. Wills Physics Laboratory, Bristol University, Tyndall Avenue, Bristol BS8 1TL, UK
‡ Dipartimento di Fisica 'E. Amaldi' Università di Roma III, Via C. Segre 2, 00154 Roma, Italy
§ Neutron Division, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire OX11 0QX, UK

LIQUID water, the medium in which life both began and persists, is in many ways a most unusual fluid. Much is known about the macroscopic properties of the condensed and gaseous states of water, but our understanding of the microscopic forces that define water structure remains incomplete[1]. This structure, described in terms of correlations between the pairs of atoms O–H, O–O and H–H (ref. 2), can be studied by neutron diffraction techniques involving isotopic substitution[3]. In particular, the signature of hydrogen bonding is apparent in neutron diffraction experiments as a peak in the pair correlation function of O and H ($g_{OH}(r)$) at about 1.9 Å (refs 3, 4). Here we extend such studies into the supercritical regime of water, in which there is no longer any distinction between the liquid and vapour phases. We find that at the supercritical temperature of 400 °C almost all hydrogen bonding is broken down, even though the thermal energy k_BT is considerably less than the energy of the hydrogen bond. Our results are markedly different from the predictions of computer simulations under comparable conditions[5] using a common model of water[5–7]. Our results provide a sensitive test of models of water structure more generally, and give some indication of the environment that solute molecules will experience in extraction and reaction processes that employ supercritical water as a solvent[8].

Attempts to describe theoretically the properties of water have generally been based on models in which the molecules are represented by an array of three or four discrete charges sited at distances corresponding roughly to the intramolecular separations. Foremost among these and originating from the seminal work of Bernal is the four point-charge ST2 model of Stillinger and Rahman[9] which reproduces many of the observed properties of real water. A subsequent and computationally more efficient model is the spherical point charge (SPC) potential, which has only one negative charge and two positive charges[5–7]. Neither of these potential models for water, nor many others, have built in the specifically directional properties characteristic of the hydrogen bond. Yet, when used in computer simulations which calculate structural and dynamical properties of water, these simple potentials produce remarkably good agreement with the measured quantities[10–12]; the hydrogen bond appears as a cooperative electrostatic phenomenon that occurs when the molecules are placed close to one another in the condensed phase. If this view of hydrogen bonding is correct then it means that much of the chemical literature, which often refers to hydrogen bonding as if it were some form of weakly covalent bond between the molecules, is seriously flawed.

To establish whether a model potential really does represent the true interaction between water molecules, we can heat water up to a sufficiently high temperature that the hydrogen bonds start to break down and investigate whether a computer simulation using this potential is able to mimic the observed structural breakdown. The process can be studied experimentally by neutron diffraction, using isotope labelling to extract the relevant correlations between O and H atoms[3], independent of OO and HH correlations. Because the coherent neutron scattering amplitudes of hydrogen and deuterium are very different, neutron diffraction experiments on mixtures of heavy and light water enable one to extract the correlation functions of interest. In particular, by focusing on the OH correlation centred at ~1.9 Å in $g_{OH}(r)$ ($g_{XY}(r)$ is a measure of the probability of finding particle X at a distance r from particle Y) one can examine how this peak changes with temperature and pressure. In a previous study[4] it was found that as the temperature is raised to 170 °C this peak diminishes a little in amplitude and moves to greater separations. Here we have extended the neutron experiments to higher temperatures and obtained structural results on H_2O in the supercritical regime.

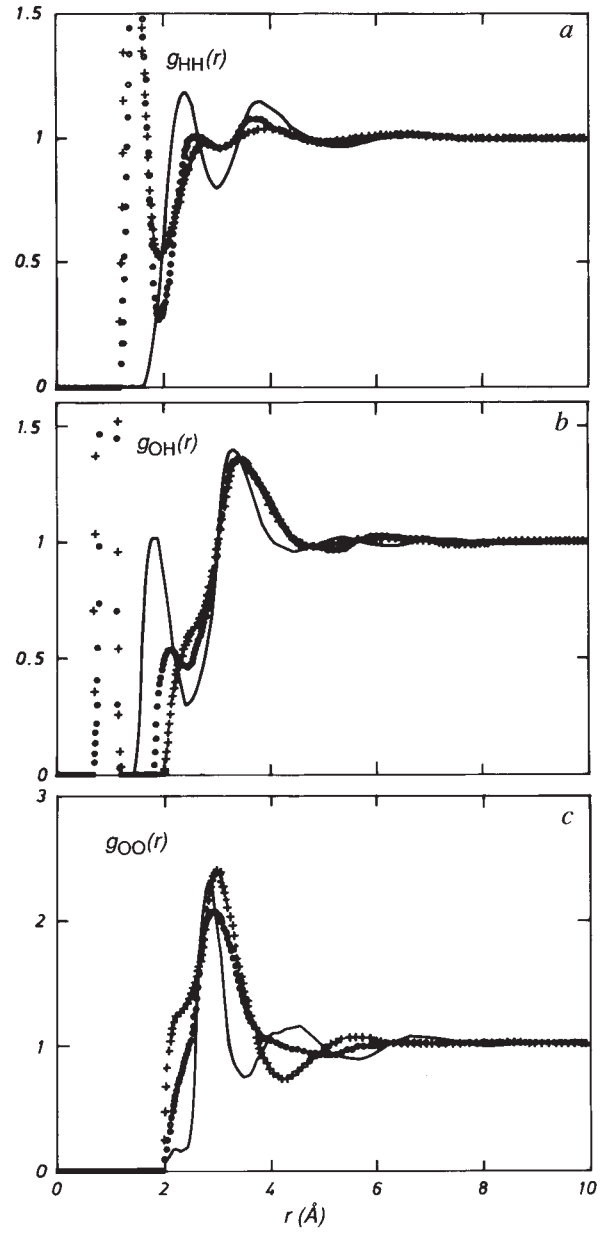


FIG. 1 The three pair radial distribution functions for H_2O at 25 °C (solid line), 300 °C (circles) and 400 °C (crosses). *a*, $g_{HH}(r)$; *b*, $g_{HO}(r)$; *c*, $g_{OO}(r)$.

Neutron diffraction with isotopic substitution (NDIS) methods are particularly well-suited to studying liquids under extreme conditions. Because neutrons are uncharged, they can

|| To whom correspondence should be addressed.

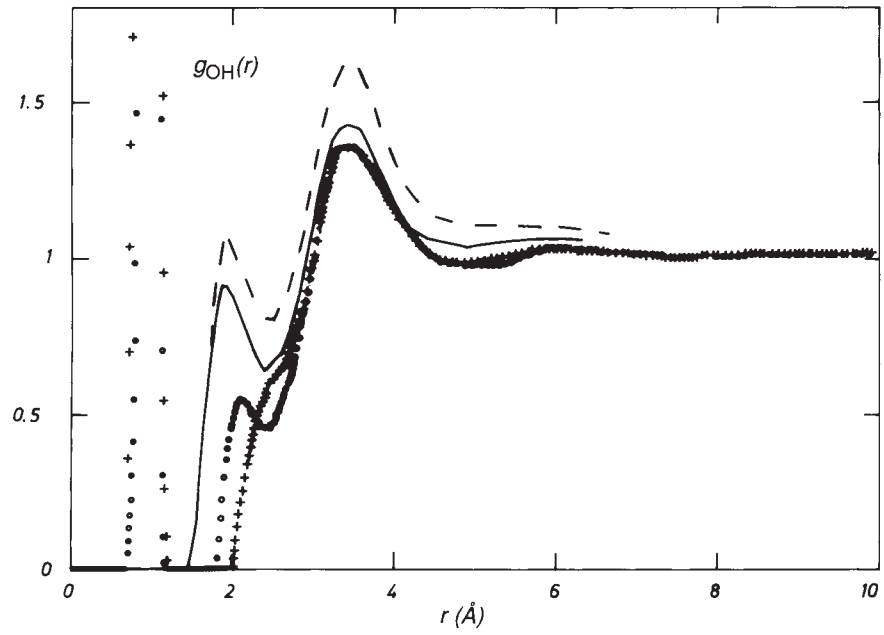


FIG. 2 Comparison of results for the oxygen-hydrogen pair radial distribution function $g_{OH}(r)$ derived from computer simulation with those determined from experiment. The simulation results correspond to SPC (see text) water at $T_r=1$, $\rho_r=1.5$ (solid line), and $T_r=1.05$, $\rho_r=1$ (dashed line) where T_r and ρ_r are ratios of the simulation temperature and density to the critical temperature and density of SPC water. The experimental results refer to $T=300$ °C$=0.88$ T_c, $\rho=0.71$ g cm$^{-3}=2.22$ ρ_c (circles) and $T=400$ °C$=1.04$ T_c, $\rho=0.64$ g cm$^{-3}=2.00$ ρ_c (crosses) where T_c and ρ_c are the experimental values. The ambient $g_{OH}(r)$ is not included.

pass through relatively large volumes of strong materials without suffering high absorption or strong scattering. Moreover, by using containers made from 'null' alloys such as titanium–zirconium ($(Ti)_{0.68}(Zr)_{0.32}$), which do not contribute to the coherent neutron scattering, one can obtain structural information that is representative of only the substances within the container[13].

Neutron diffraction experiments were carried out on the SANDALS diffractometer located on the ISIS spallation neutron source at the Rutherford Appleton Laboratory. This instrument is optimized for diffraction studies of systems that contain light atoms, and has demonstrated that NDIS experiments based on H/D substitution can be used to provide valuable quantitative information on the structure of hydrogenous materials[14]. Data were obtained on three isotopic mixtures of water (H_2O) and heavy water (D_2O): pure H_2O, $(H_2O)_{0.5}(D_2O)_{0.5}$ and $(H_2O)_{0.003}(D_2O)_{0.997}$. Experiments were carried out for each sample at two conditions: temperature $T=300(1)$ °C, pressure $p=95(5)$ bar, density $\rho=0.72$ g cm^{-3} ($n=0.072$ atoms Å^{-3}) and $T=400(1)$ °C, $p=800(10)$, bar, $\rho=0.66$ g cm^{-3} ($n=0.066$ atoms Å^{-3}, where n is the total number density of atoms in the fluid). (Recall for H_2O the critical point occurs at $T_c\sim374$ °C, $p_c\sim221$ bar, $\rho_c\sim0.32$ g cm^{-3}.) Additional data were obtained for the empty container, the background and a vanadium plate. The data were analysed by standard methods[15] which included corrections for sample and container absorption, multiple scattering and normalized to a vanadium standard plate.

On the justifiable assumption that the structure of the fluid is independent of isotopic state, the three pair radial distribution functions were calculated from the diffraction data of the three isotopically distinct samples. It is worth noting that $g_{OO}(r)$ is the least well-determined function in the experiment, because (1) the isotopic substitution was made on the protons, and so the difference between neutron diffraction patterns is entirely due to the different scattering properties of H and D nuclei; and (2) the concentration of oxygen is half that of hydrogen. As a result, $g_{OO}(r)$ contributes only ~10% to the total scattering pattern in two of the three experiments.

As might be anticipated, there are some obvious differences in the $g_{ij}(r)$'s at ambient temperature, 300 °C and 400 °C (Fig. 1). The structure of the H_2O molecule itself remains essentially unchanged over the temperature range of the experiments, with its O–H and H–H intramolecular peaks centred near 1 Å and 1.55 Å respectively. The fact that integration of $4\pi n\, r^2 g(r)$ over the range of these peaks gives values for the coordination numbers $\bar{n}_O^H=0.98$ (5) and $\bar{n}_H^H=2.15$ (20)—within error, the correct stoichiometry for H_2O—confirms that the data have been properly normalized.

In the intermolecular region, significant variations are observed in $g_{OH}(r)$ and $g_{HH}(r)$. Arguably the most interesting variation with temperature is in the H-bonding regime, the signature of which is the peak located at ~1.9 Å in $g_{OH}(r)$. This peak, which is pronounced at room temperature, is affected dramatically by an increase in temperature: at 300 °C the peak has shifted to ~2.1 Å and is ~1/3 the height of that at ambient conditions. At 400 °C the peak has completely disappeared, suggesting that the H-bonding between molecules has been largely destroyed at this temperature. The coordination number of H atoms around a central O atom, calculated by integrating over the first intermolecular peak, is 0.5 ± 0.2 at 300 °C and is negligible at 400 °C. This contrasts with a value at ambient temperature of 1.6–1.9, suggesting that the degree of H-bonding at 300 °C is ~25% of that in room-temperature water.

The results also show an appreciable drop in intensity in $g_{HH}(r)$ at ~2.3 Å, corresponding to the nearest-neighbour intermolecular peak also associated with hydrogen bonding. Interestingly, as the temperature increases, this peak becomes less intense and shifts from ~2.6 Å to 2.8 Å and is probably related to the weakening of orientational correlations between nearest-neighbour water molecules. The form of $g_{OH}(r)$ beyond the first coordination shell appears to be similar at 300 °C and 400 °C, but appreciably different from the more structured room-temperature results in this region. The $g_{HH}(r)$ function also becomes less structured at the higher temperatures.

These results are initially surprising when one considers simply scaling the energies by k_BT (where k_B is Boltzmann's constant). For example, at 400 °C the molecular kinetic energy of H_2O is still 30% of the apparent H-bond energy of 24 kJ mol^{-1} (ref. 16). But any detailed statistical analysis must take account of the states available to unbound water molecules, and it is possible that a more sophisticated mean-field theoretical approach will explain why H-bonding disappears at relatively low temperature.

As mentioned above, $g_{OO}(r)$ is less accurately determined than $g_{OH}(r)$ and $g_{HH}(r)$. But as it has been calculated in the same manner at all three temperatures, the results for $g_{OO}(r)$ can be compared in a semi-quantitative way. Most obvious is a significant broadening of the main peak. There is also a commensurate increase in coordination number for nearest-neighbour water

molecules which indicates that the water substance is behaving more like a normal liquid[17].

We now consider information from the recent computer simulation study by Cummings and co-workers[5] of SPC water near its critical point, which is calculated to be at to $T_c \sim 578$ K, $\rho_c \sim 0.270$ g cm^{-3}. These authors were particularly interested in examining SPC water structure around the critical point and carried out two molecular-dynamics calculations (Fig. 2). Although our experiments were carried out at different reduced temperatures and densities, the $g_{ij}(r)$'s are demonstrably different from those of the simulation (Fig. 2). This is most visible in the region ~2 Å where the experimental $g_{OH}(r)$ is significantly changed by temperature. The important point to note is that the simulations were carried out at lower reduced densities than in our experiment and yet still indicate that strong orientational correlations exist between near-neighbour model water molecules close to the critical point. In contrast, our experimental results shows that such correlations are already absent for densities which are twice the critical density, that is for states which one might describe as liquid-like.

Our study provides a benchmark for future studies of water and ionic solutions under supercritical conditions. The level of detail of the structural information should enable theoreticians to adjust their models to correlate better with experimental results. □

Received 26 August; accepted 2 November 1993.

1. Neilson, G. W. & Enderby, J. E. (eds) *Water and Aqueous Solutions* (Adam Hilger, Bristol, 1986).
2. Soper, A. K. & Phillips, M. G. *Chem. Phys.* **107,** 47–60 (1986).
3. Enderby, J. E. & Neilson, G. W. in *Water, A Comprehensive Treatise* Vol. 1 (ed. Franks, F.) Ch. 1 (Plenum, New York, 1979).
4. Yamaguchi, T. & Soper, A. K. ISIS Annual Report No. A110, Report No. 91–050, Vol. 2 (Rutherford Appleton Laboratory, Chilton, 1991).
5. Cummings, P. T., Cochran, H. D., Samonson, J. M., Mesmer, R. E. & Karaboni, S. *J. Chem. Phys.* **94,** 5606–5621 (1991).
6. Mountain, R. D., *J. chem. Phys.* **90,** 1866–1870 (1989).
7. Guissani, Y. & Guillot, B. *J. chem. Phys.* **98,** 8221–8235 (1993).
8. Pichal, M. & Sifner, O. (eds) *Properties of Water and Steam* (Hemisphere, New York, 1989).
9. Stillinger, F. H. & Rahman, A. *J. chem. Phys.* **60,** 1545–1557 (1974).
10. Corongiu, G. & Clementi, E. *J. chem. Phys.* **97,** 2030–2038 (1992); **98,** 2241–2249 (1993).
11. Impey, R. W., Klein, M. C. & McDonald, I. R. *J. chem. Phys.* **74,** 647–652 (1981).
12. Ferrario, M. & Tani, A. *Chem. Phys. Lett.* **121,** 182–186 (1985).
13. Neilson, G. W., Broadbent, R., Howe, M. A. & Frank, E. U. in *Properties of Water and Steam* (eds Pichal, M. & Sifner, O.) 335–346 (Hemisphere, New York, 1989).
14. Soper, A. K. & Turner, J. *Int. J. mod. Phys.* **7,** 3049–3076 (1993).
15. Soper, A. K., Howells, W. S. & Hannon, A. C. Report No. 89–046 (Rutherford Appleton Laboratory, Chilton, 1989).
16. Walrafen, G. E., Fisher, M. R., Hokmabadi, M. S. & Yang, W-H., *J. chem. Phys.* **85,** 6970–6982 (1986).
17. Luzar, A. *J. chem. Phys.* **91,** 3603–3613 (1989).

ACKNOWLEDGEMENTS. We thank R. Evans for discussions in the course of preparing this manuscript. We also thank R. Done and other members of ISIS sample environment group for technical assistance. This work was supported by the UK SERC and the Italian CNR.

Potential effects of cloud optical thickness on climate warming

George Tselioudis*†, Andrew A. Lacis*, David Rind* & William B. Rossow*

* NASA Goddard Institute for Space Studies, 2880 Broadway, New York, New York 10025, USA
† Columbia University, Morningside Heights, New York, New York 10027, USA

CLIMATE warming can cause changes in the optical properties of low clouds, which may in turn amplify or diminish the warming[1,2]. But both the sign and magnitude of such feedbacks have been uncertain, largely because the observational evidence for variations in the large-scale optical properties of clouds has been very limited. Recently, analysis of data from the International Satellite Cloud Climatology Project yielded a relationship between low-cloud optical thickness and cloud temperature that implies a positive feedback between clouds and climate[3]. Here we use a two-dimensional radiative–convective model to assess the effect of such a feedback on the climate change associated with a doubling of the atmospheric carbon dioxide concentration. We find that, zonally averaged, the feedback is positive in the Northern Hemisphere and is stronger at lower than at higher latitudes. The positive feedback amplifies the overall global climate sensitivity, and the latitudinal gradient in the strength of the feedback acts to eliminate the high-latitude amplification of the greenhouse warming predicted by most climate models.

The radiative effects of clouds on climate are determined by the relative strengths of opposing effects; the solar albedo (cooling effect) and the thermal greenhouse (warming effect). Averaged over an annual cycle and the whole globe, clouds produce a net cooling effect due to the predominance of their solar albedo effect over their thermal greenhouse contributions[4–6]. But the important question of how clouds will actually change as the climate warms owing to trace gas emissions is complex and remains largely unanswered.

Whether cloud changes will produce a positive or negative feedback depends on the detailed changes of cloud radiative properties: mainly cloud cover, cloud height and cloud optical thickness. In an experiment with the Goddard Institute for Space Studies (GISS) general circulation model[7] where the atmospheric CO_2 concentration was doubled, clouds were found to produce a positive feedback due primarily to a decrease in cloud cover (mostly low clouds) and an increase in cloud height (low clouds replaced by cirrus). Cloud optical thickness was prescribed in the experiment to depend on cloud height and vertical extent. Therefore, no direct cloud optical-thickness feedback was allowed to operate.

Cloud optical property feedbacks are difficult to determine, partly because the influences of dynamic and microphysical processes on cloud water content are not fully understood, and also because large-scale measurements of cloud optical thickness have not been available. A relationship between cloud water content and temperature (derived from limited mid-latitude continental observations and applied to a one-dimensional model[18]) has suggested a negative cloud optical-thickness feedback on climate warming. On the other hand, GCM studies with schemes to predict cloud optical properties[9–11], obtained strong cloud optical-thickness feedbacks that changed in both magnitude and sign when different parameters were used for cloud processes.

Global observations of cloud optical thickness have recently become available in the International Satellite Cloud Climatology Project (ISCCP) dataset[12], which contains detailed information on the global distribution of cloud radiative properties and their diurnal and seasonal variations, as well as correlative information on the vertical distribution of temperature and humidity in the troposphere. An analysis of one year of ISCCP data[3] shows that changes in the optical thickness of Northern Hemisphere low clouds are correlated with changes in the mean cloud temperature on different time and space scales. A parameter f, defined as the normalized change of cloud optical thickness with cloud temperature,

$$f = \frac{1}{\tau}\frac{d\tau}{dT} \qquad (1)$$

was used to describe this relation between optical thickness and temperature.

The latitudinal and seasonal variations of this relation are shown in Fig. 1, where f for low clouds between 15° and 55° N is plotted for the four seasons separately over land (Fig. 1*a*) and ocean (Fig. 1*b*). (Tropical clouds are excluded from the analysis

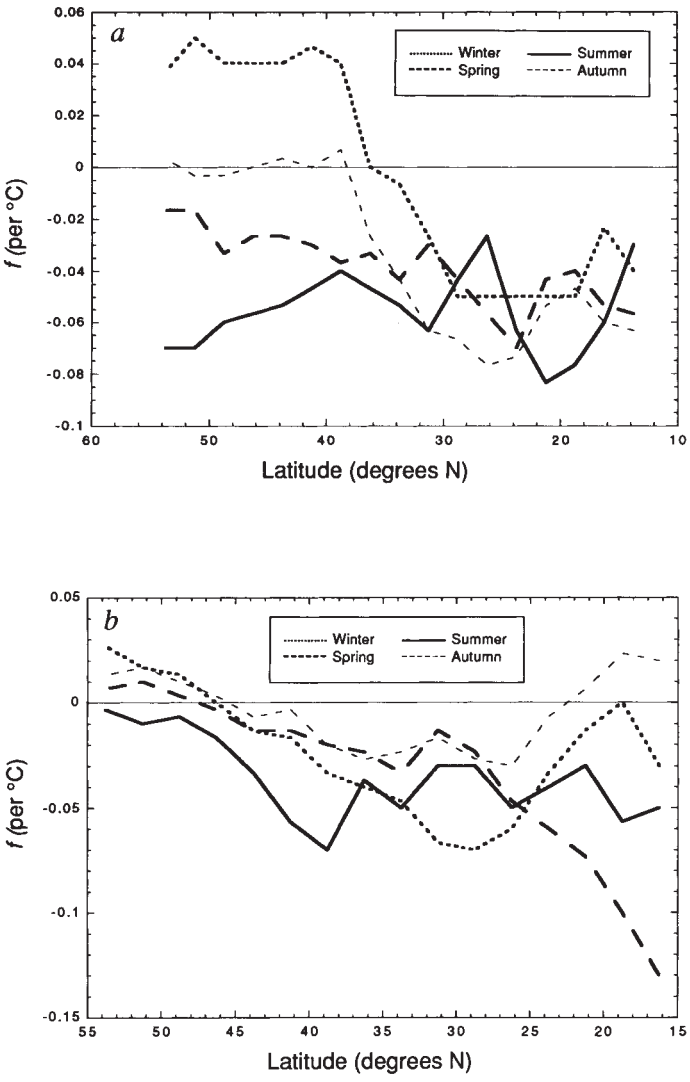


FIG. 1 Latitudinal variation of f (see equation (1)) in the range 15–55° N for low clouds over land (*a*) and low clouds over ocean (*b*) in Northern Hemisphere winter, spring, summer and autumn. The points are plotted at the ISCCP resolution of 2.5°.

because the yearly temperature variation in the tropics is too small to obtain meaningful statistics; polar clouds are excluded because of the large uncertainties in the retrieval of cloud optical thickness over ice/snow-covered areas.) It can be seen that, for winter continental clouds, optical thickness increases with temperature, as f is positive with an average value of ~0.045. This value is consistent both with the temperature variation of the adiabatic cloud water content[13] and with the observations discussed in ref. 8. For warmer continental and almost all maritime clouds, however, cloud optical thickness decreases with temperature with an average value for f between −0.04 and −0.05. The only notable exception is subtropical maritime clouds in autumn.

The latitudinal and seasonal variations of f (Fig. 1) show an overall pattern of increasing optical thickness with temperature in colder cloud ensembles, and decreasing optical thickness with temperature in warmer cloud ensembles. Additional analysis of the ISCCP data[3] showed that this characteristic pattern remains consistent when clouds at the same latitude and season, sorted into cold and warm ensembles by the day-to-day variations of the temperature field, are examined. Further, the relation between optical thickness and temperature is not significantly different in large-scale dynamical regimes defined by different large-scale vertical velocities[14], although the mean optical thickness does appear to be systematically different. This connection between f and temperature points towards changes in temperature-dependent cloud process(es) as the primary explanation for the relation. G.T. *et al.*[3] suggest, but do not establish, that the main reason for the transition to negative f values at warm temperatures is an increase with temperature of the efficiency of precipitation relative to condensation. But the exception of the autumnal subtropical clouds over the ocean implies that dynamical processes may also play an important role.

Figure 2*a* shows the annual mean f in the 15–55° N latitude range for low clouds over land, ocean, and for the zonal mean cloud field. (Annual mean values of f represent averages over the seasons weighted by the cloud amount and the solar insolation in each season.) The zonal/annual mean values of f are negative everywhere, with high-latitude values ~−0.02 and low-latitude values ~−0.045. The curves for continental and maritime clouds are quite similar at the lower (warmer) latitudes of the range but differ at the higher (colder) latitudes, where the maritime values of f are significantly more positive than the continental ones. This difference is due to the low values of f observed in summer clouds over land (Fig. 1*a*). Averaging over all seasons smoothes the 'anomalous' behaviour of autumnal subtropical clouds over ocean and, more importantly, hides the seasonal differences in mid-latitude clouds over land.

The annual mean curves shown in Fig. 2*a* suggests a possible cloud optical-thickness feedback that varies with latitude and that, at the higher latitudes of the range, can also vary with location of the clouds over land or ocean. The analysis that follows is meant to illustrate the significance of the latitudinal and longitudinal variations of the potential feedback by assessing the annual mean surface temperature response to changes in the annual mean low-cloud optical thickness.

We use a two-dimensional radiative–convective dynamic equilibrium model, with 9 levels in the vertical and 24 latitude intervals from pole to pole. The meridional transports of sensible heat, latent heat and geopotential energy are specified from the control run of the 8°x10° version of the GISS GCM[15]. Latitude-dependent annual average profiles of atmospheric gases and aerosols and the surface properties of the model are also taken from the GISS GCM control run, but the annual mean cloud cover, optical thickness and vertical distribution at each latitude are taken directly from the ISCCP dataset[12].

The method used to perform the feedback analysis is adapted from Hansen *et al.*[7]. The two-dimensional model is first run to equilibrium to establish the reference surface temperatures. The optical thicknesses of all low clouds are then reduced by 50%, and the model is again run to equilibrium with no other structural or feedback changes allowed to operate. The resulting change in equilibrium surface temperature (ΔT_f) can then be calibrated for specific changes in low-cloud optical thickness. For this 50% reduction in low-cloud optical thickness, the equilibrium surface temperature increases by amounts that vary from 1.5 °C in the tropics to 2 °C in the mid-latitudes.

The values of f at each latitude, derived from the observed cloud changes (Fig. 2*a*), are then used to calculate the temperature change (ΔT_t) that would be needed to produce a 50% decrease in low-cloud optical thickness. For a 50% decrease in τ, equation (1) gives $\Delta T_t = -\ln 2/f$. The ratio of the temperature change (ΔT_f) resulting from the cloud optical-thickness decrease to the temperature change (ΔT_t) needed to produce that same decrease determines the feedback efficiency G, where $G = \Delta T_f/\Delta T_t$. Knowing the feedback efficiency for low-cloud optical-thickness changes at each latitude, we evaluate the net effect of this feedback in a climate change scenario. If ΔT_0 is the surface temperature change produced by doubling the atmospheric CO_2 concentration when no feedbacks occur, then the equilibrium temperature change if all feedbacks are included would be $\Delta T_{eq} = \Delta T_0/1-(G+g)$, where G is the feedback efficiency for low-cloud optical-thickness changes and g is the sum

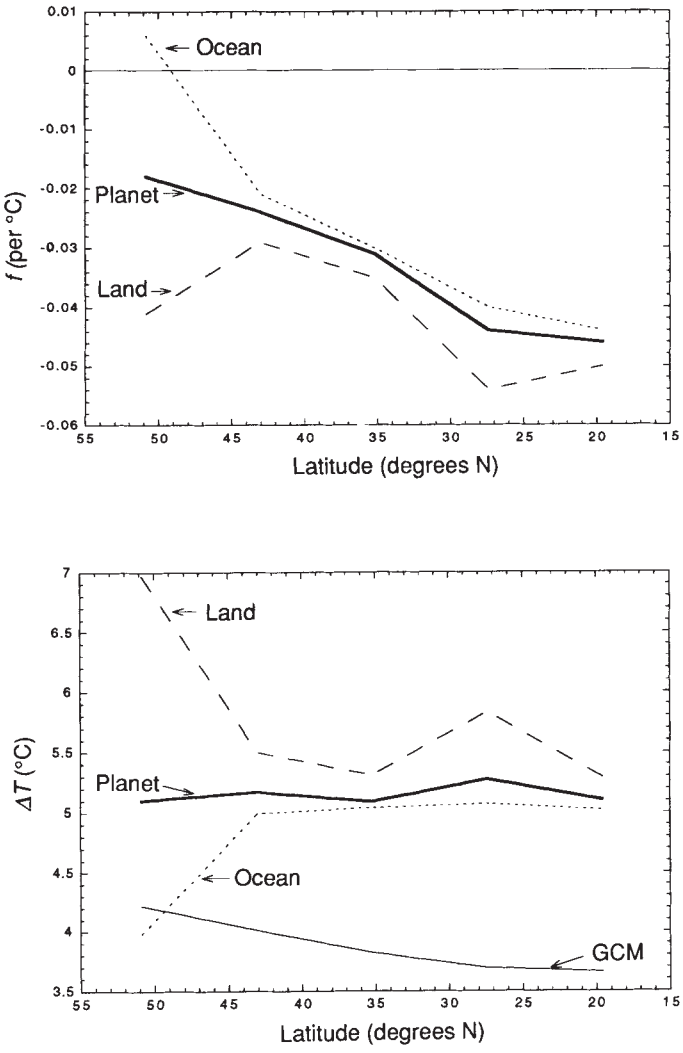


FIG. 2 *a*, Latitudinal variation of the annual mean value of *f* for low clouds over land (dashed line), low clouds over ocean (dotted line) and the zonal mean low clouds (solid line) in the range 15–55° N. The points are plotted at the model resolution of 8°. *b*, Change in surface temperature (ΔT) from the $2 \times CO_2$ run of the GISS GCM (thin solid line) and estimated change of the same quantity with the low-cloud optical-thickness feedback included, for a land-covered planet (dashed line), an ocean-covered planet (dotted line) and the Earth (thick solid line).

of the feedback efficiencies for all other feedback processes operating in the GISS GCM $2 \times CO_2$ experiment[7] (the sum of the feedback efficiencies includes contributions from changes in the amount and distribution of water vapour, the atmospheric lapse rate, the ground albedo and the cloud height and cloud cover). Knowing g and ΔT_0 from the GISS GCM $2 \times CO_2$ experiment, we then calculate an estimate for the new equilibrium temperature ΔT_{eq} that would be obtained with the low-cloud optical-thickness feedback also included.

The feedback analysis is done separately for continental and maritime clouds, as well as for the zonal mean cloud field. In other words, the equilibrium temperature change is calculated for an ocean-covered planet, a land-covered planet and a planet with the actual continental/maritime low-cloud ratio of the Earth. Figure 2*b* shows the change in surface temperature from the $2 \times CO_2$ simulation of the GISS GCM with no direct cloud optical-property feedbacks included and the estimated change in surface temperature with the low-cloud optical-thickness feedback included for the three cases mentioned above. For the zonal mean cloud field, the additional cloud feedback is positive, increasing the greenhouse warming at all latitudes by amounts that vary from 1.5 °C in the tropics to 0.5 °C in mid-latitudes. This latitudinal variation in the values of the optical thickness feedback efficiency eliminates the high-latitude amplification of the predicted warming. The 'continental' and 'maritime' temperature changes are relatively similar at lower latitudes, with continental clouds producing larger increases than maritime ones. At higher latitudes, however, a significant difference occurs. Maritime clouds produce smaller temperature increases (and even a small decrease at 52° N) resulting in a reduction of the greenhouse warming at that latitude. Continental clouds, on the other hand, produce a significant temperature increase that reaches 1.8 °C at the highest latitude of the range.

These results provide a new and complex perspective on the question of cloud optical-property feedbacks in climate. The large difference in the high-latitude warming between the continental and maritime cloud cases (Fig. 2*b*) suggests that the proposed cloud feedback may introduce longitudinal temperature contrasts. Furthermore, the large high-latitude warming in the continental cloud case (Fig. 2*b*), which is due to the very low values of *f* in the summer (Fig. 1*a*), implies that the latitudinal structure of the greenhouse warming can vary significantly with season. Both issues must be examined further through the use of fully interactive three-dimensional GCMs. When the annual mean relation between optical thickness and temperature (weighted by seasonal cloud amount and solar insolation) is used in a two-dimensional model, it produces a positive climate feedback everywhere, but one that is stronger at low, rather than at high latitudes. This acts to reduce, or even to eliminate, the high-latitude amplification characteristic of model simulations of the greenhouse warming. Such a reduction, as it affects the equator-to-pole temperature gradient, could also affect changes in the strength of the atmospheric circulation in a $2 \times CO_2$ scenario, which would feed back on water vapour, oceanic changes and the cloud field itself. The differences between 'continental' and 'maritime' cloud regimes imply the possibility of longitudinal gradients in the magnitude of the greenhouse warming, which could alter the strength of the eddy component of the circulation.

This study illustrates the possibility of an important feedback, the nature of which may be different from what has been believed previously. One clear implication is that a detailed knowledge of the latitudinal, seasonal and regional variations of cloud feedback is necessary in order to determine the actual role of clouds in global climate change. The process of cloud feedback is influenced by a number of components, one of which is the change in optical thickness with temperature. Our work shows that the cloud optical-thickness feedback has a notable latitude dependence that tends to counteract the high-latitude amplification of the warming predicted by most climate models for doubled atmospheric CO_2 concentrations. Understanding the full impact of cloud changes on climate requires determination of the other feedback components as well. □

Received 23 June; accepted 5 November 1993.

1. Wang, W.-C., Rossow, W. B., Yao, M. S. & Wolfson, M. *J. atmos. Sci* **38,** 1167–1178 (1981).
2. Platt, C. M. R. & Harshvardhan *J. geophys. Res.* **93,** 11051–11058 (1988).
3. Tselioudis, G., Rossow, W. B. & Rind, D. *J. Clim.* **5,** 1484–1495 (1992).
4. Ellis, J. S. thesis, Colorado State Univ. (1978).
5. Ramanathan, V. *et al. Science* **243,** 57–63 (1989).
6. Hartmann, D. L., Ockert-Bell, M. E. & Ebert, E. E. *J. Clim.* **5,** 1281–1304 (1992).
7. Hansen, J. *et al.* in *Climate Sensitivity: Analysis of Feedback Mechanisms* (eds Hansen, J. E. & Takahashi, T.) 130–163 (Geophys. Monogr. Ser. No. 29, Am. Geophys. Un., Washington DC, 1984).
8. Somerville, R. C. J. & Remer, L. A. *J. geophys. Res.* **89,** 9668–9672 (1984).
9. Mitchell, J. F. B., Senior, C. A. & Ingram, W. J. *Nature* **34,** 132–134 (1989).
10. Roeckner, E. *Nature* **335,** 304 (1988).
11. Le Treut, H. & Li., Z.-X. *Climate Dyn.* **5,** 175–187 (1991).
12. Rossow, W. B. & Schiffer, R. A. *Bull. Am met. Soc.* **72,** 2–20 (1991).
13. Betts, A. K. & Harshvardhan. *J. geophys. Res.* **92,** 8483–8485 (1987). **(see ref. 2)**
14. Tselioudis, G. thesis, Columbia Univ. New York (1992).
15. Hansen, J. *et al.* Mon. Weath. Rev. **111,** 609–662 (1983).

ACKNOWLEDGEMENTS. We thank M. Sato for assistance with the model and P. Stone and T. Del Genio for discussions. This work was supported by the NASA Climate Program.

Evidence from the Oman ophiolite for sudden stress changes during melt injection at oceanic spreading centres

Benoît Ildefonse, Adolphe Nicolas & Françoise Boudier

Laboratoire de Tectonophysique, Université Montpellier II-CNRS, 34095 Montpellier cedex 05, France

THE system of dikes in the uppermost mantle below oceanic spreading centres bears witness to the processes involved in the supply of melt to the ridge. Such systems cannot be studied on currently active ridges, but can be seen in ophiolites; in particular, the Oman ophiolite allows us to examine segments of a fast-spreading ridge[1]. Here we present a statistical analysis of the orientation of intrusions in the uppermost mantle section of the Oman ophiolite spreading centres, which points to the existence of two systems: dikes whose azimuth is parallel to the sheeted dikes (which we thus consider as related to the extensional lithospheric stress field) and sills, roughly parallel to the Moho[2]. Branching contemporaneous dikes and sills are frequently observed. We suggest that these relations can be explained by repeated sudden changes of the dominant stress field from lithospheric to asthenospheric. Such changes could result from the episodic relaxation of lithospheric stress in response to tension fracturing of the lithospheric lid, related to sudden melt surges. This episodic behaviour is recorded in the sheeted dike complex, with each dike (each about one metre wide) representing one of these fractures.

Various types of dikes and veins may be distinguished and mapped separately in Oman[2]. The density of dikes in mantle sections of this ophiolite is variable, but it is always high in diapiric areas. In those diapirs, it is still possible to observe the original dike organization, in contrast to other areas where mantle flow may have deformed and transposed the dikes parallel to the flat-lying foliation. We can separate the observed intrusions into two main groups[2]. The first group comprises early 'indigenous' gabbro and pyroxenite dikes (equilibration temperature >1,200 °C) which intrude into melt-impregnated peridotites and are bounded by dunite walls (formed by wall-rock reaction of the circulating melt in the dike[3]). Dunite veins belong to this group. Pyroxenite dikes are dominantly olivine-bearing websterites or clinopyroxenites; they may grade into impregnation clots or into chromite schlieren and pods. The second group of gabbro, pyroxenite and diabase dikes is called 'intrusive' (emplaced at temperatures below the peridotite solidus) because the dikes have sharp and straight walls without reaction with the surrounding peridotites. Overall, this group represents a continuum in composition, ranging from pegmatitic gabbros and pyroxenites displaying comb structures (induced by the crystallization of melt from the cooler walls toward the core of the dike) to diabase dikes with chilled margins, the latter corresponding to host-rock temperatures <450 °C (ref. 4). Thickness ranges from a few centimetres to 1 m in both groups.

Whatever the nature and the time of intrusion of the dikes, analysis of their orientations in diapiric areas shows that they tend to cluster in two groups as already shown in two massifs of southern Oman (Fig. 1)[2]. We present here new data from the Batin area (Fig. 1, Southern Wadi Tayin massif), where a zone of steeply plunging foliations and lineations has been identified in the mantle section, suggesting that it is also a diapiric structure. In this area, after rotation of all measurements into the palaeo-ridge reference frame (Moho rotated to the horizontal, ridge trend defined by diabase sheeted-dike trend), the intrusions group into two main orientations which, admittedly, are not always clear-cut (Fig. 2). One set of dikes is nearly vertical and sub-parallel to the sheeted-dike trend in this region, the other forms sills and moderately inclined dikes which have a tendency to wrap round the presumed mantle diapir (Fig. 2). Another common observation is that cross-cutting dikes and sills may not display any evidence of relative chronology. The dikes can post-date the sills, or vice-versa. We can also observe branching dikes and sills in which melt has been flowing from one into the other (Fig. 3). The partitioning in dikes and sills is generally stronger in the intrusive group than in the indigenous group, where dikes and sills often display more dispersed preferred orientations. On average, there are more sills in the indigenous group and more vertical dikes in the intrusive group[2].

It has been proposed previously that in peridotites, as in other formations, dike orientations are controlled by the ambient stress field, the dike being perpendicular to the σ_3 principal minor stress axis[1,5,6]. Here, the group of vertical dikes sub-parallel to the sheeted-dike average orientation (assumed to represent the ridge axis), would be thus related to lithospheric fracturing, that is a stress field with σ_3 horizontal and perpendicular to the ridge. This implies that the lithospheric stresses are transmitted, at least partially, to the upper mantle beneath the ridge. On the other hand, the occurrence of sills and gently dipping dikes implies, at least locally, the existence of another stress field, with σ_3 nearly vertical. The switch between the two inferred stress fields must be rapid, both in time and space, because of the mutual cross-cutting and branching figures mentioned above. We believe it is best explained by a sudden and episodic relaxation of the lithospheric tensional stress, as demonstrated in similar context at the divergent plate boundary in northeast Iceland[7]. Such an event may be related to the release of melt in the crust. It has been proposed that the hydrofracturing events responsible for melt rise are episodic and short-lived, being recorded best in the sheeted-dike complex. Each dike, 1 m wide on average, represents a single and rapid melt surge (a dike stays molten for at most a few weeks). Consequently, melt surge through hydrofractures brings to the accreting crust a quantity of melt large enough to produce, on average, 1 m of crust[1]. We speculate that this rapid opening would relax the lithospheric stress and

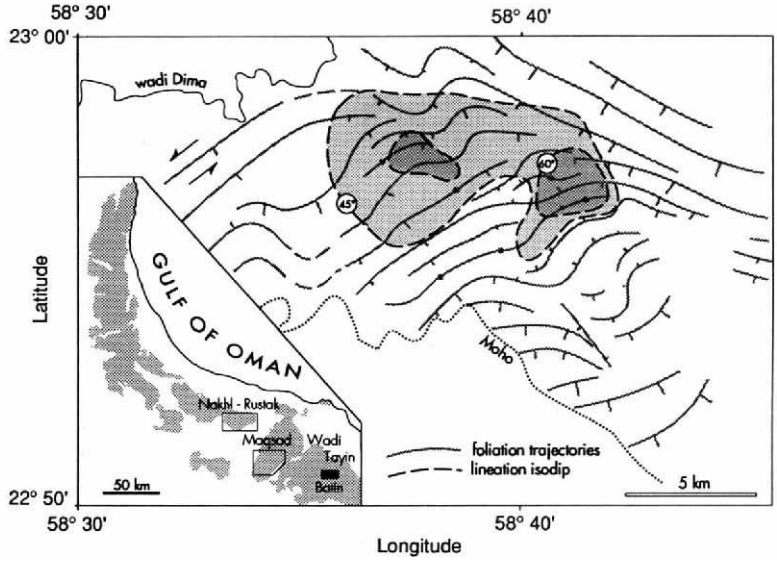


FIG. 1 Simplified structural map of the mantle section in the Batin area. The foliation trajectories fit measurements collected over 220 stations. Grey-filled contours of the lineation isodip locate the presumed diapiric centre. The location of the two previously studied massifs (Nakhl-Rustak and Maqsad)[2] is shown in the inset map.

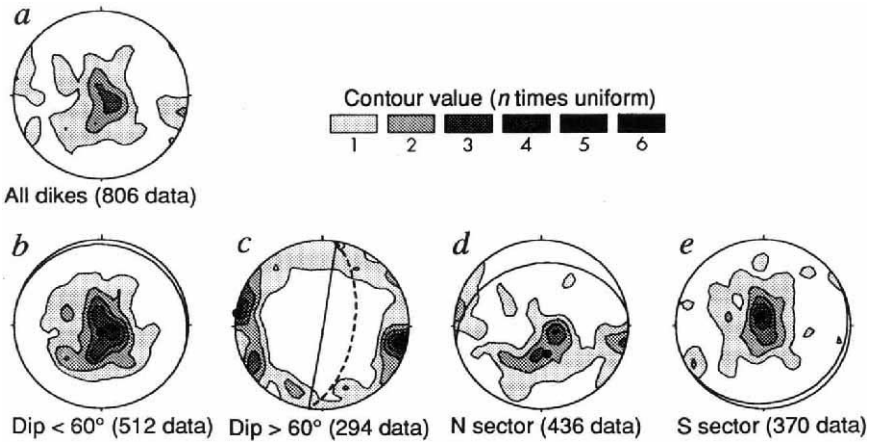


FIG. 2 Lower-hemisphere stereographic projections of the dike measurements (poles of the planes). All data have been rotated 30° to the north about a nearly east–west (N100) horizontal axis, that is the rotation inferred from the cross-sections, in order to restore the Moho to a horizontal position. Great circles (solid and, where appropriate, dashed lines) represent the mean planes in (*b*) to (*e*). *a*, Bulk data (pyroxenite and gabbro intrusions) showing the partition into sills and dikes. This data may be divided into two sets, one (*b*) for moderately inclined (dip < 60°) dikes and sills, and another (*c*) for strongly inclined (dip > 60°) dikes. In *b*, the calculated mean plane is nearly flat-lying; in *c*, the calculated mean plane is vertical, and sub-parallel to the average sheeted-dike orientation in this area (great circle line, dashed). The data may also be divided into two geographical sets, north and south of the centre of the diapir. In the northern sector (*d*), the sills and moderately inclined dikes dip, on average, to the north; in the southern sector (*e*) they are on average flat-lying, dipping slightly to the south.

momentarily allow some other process to cause the switch of the principal stress axes. Such a switch could result from the dike intrusions themselves. As suggested earlier[8], providing the magma supply is large enough, dike intrusion can locally increase the least compressive σ_3 until it is no longer the least compressive stress. If the magma supply does not shut off at this point, then subsequent sheet intrusions should be orthogonal to the 'parent' dike, taking the form of either ridge–perpendicular dikes or sills. We observe both dikes and sills, but sills are predominant (Fig. 2). The presence of the Earth's surface acts to favour the development of sills, because they produce smaller stress changes than ridge-perpendicular dikes[8]. But, we suggest that the abundance of sills may also be explained by the following model, which takes into account the fundamentally dynamic character of the sub-ridge asthenosphere, as inferred from the structural mapping of Oman peridotites[1]. Two-dimensional models of asthenospheric flow below fast-spreading ridges show a narrow sub-ridge mantle upwelling[9–11] which implies the occurrence of a fast divergent non-coaxial flow away from the diapir axis, immediately below the Moho[9,12]. Such a forced flow should produce an asthenospheric stress field such that the stress axis σ_1 (which is vertical above the diapir axis) rotates rapidly towards the horizontal position at the edges of the diapir and finally dips away from the ridge[9]. Thus, at the top of the diapiric structure, slightly off-axis and below the Moho, one must expect a stress field with a steep σ_3 and the (σ_1, σ_2) plane flat-lying or moderately dipping away from the centre of the diapir. We suggest that the sill injection is controlled by this 'asthenopheric' stress field related to mantle diapirism (Fig. 4). This hypothesis is supported by the predominance of sills in the indigenous (high-temperature, asthenospheric) group and by the tendency of the sills to dip away from the diapiric area[2] (Fig. 2*d* and *e*), although it is difficult to locate precisely the diapir axis because of the complex geometry. The two proposed explanations, however,

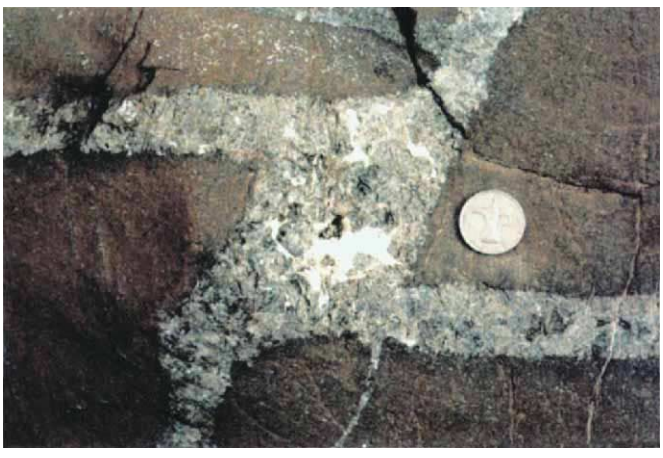


FIG. 3 Branching dike and sill of plagioclase-bearing pyroxenite. The coin has a diameter of 28 mm.

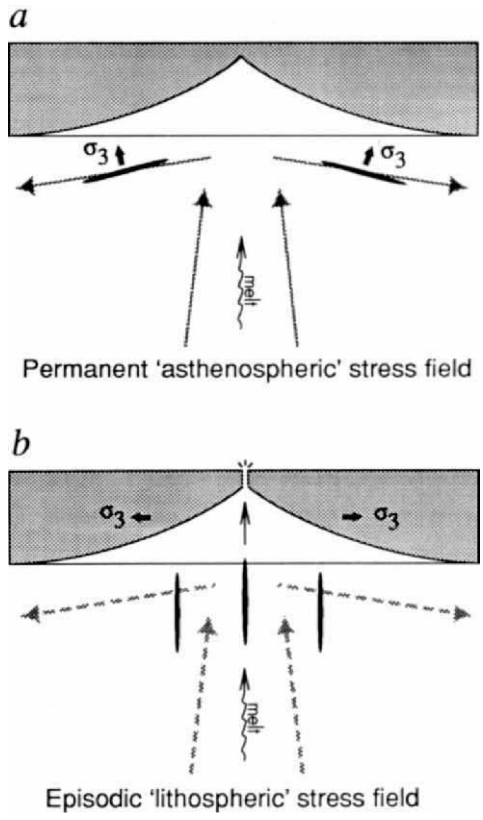


FIG. 4 Schematic cross-sections of the 'breathing ridge' model perpendicular to the ridge axis. The permanent stress field related to the diverging asthenospheric flow (*a*) allows the sills and moderately inclined dikes to open when the lithospheric tensional stress field is relaxed. In (*b*), the lithospheric tensional stress becomes progressively dominant, allowing the vertical dikes to open, until the next intrusion-related stress relaxation.

are not incompatible and may operate together to allow the rapid switch between dike and sill intrusions.

One of the goals of sub-ridge asthenopheric models is to explain how melt is focused at the ridge[10]. The sheet-intrusion net described here may be part of the explanation, with sills and ridge-perpendicular dikes capturing flow from ridge-parallel dikes, and moving it to the ridge axis. One could propose, to explain the observed system, a simple elastic model in which a local switch between the principal stresses would be related only to dike intrusion itself. We believe (on the basis of the field arguments) that it is better explained by including in the model the effect on the stress field of plastic flow in the upper mantle. Such a 'breathing ridge' model implies that at the onset of the melt surge, even at Moho depth beneath the ridge axis, the regional lithospheric stress field can temporarily predominate over the more local asthenospheric stress field that is related to the diapir uprise. This model explains how, during a single magmatic event, lithospheric and asthenospheric control may be exerted successively, allowing dikes and sills to be opened almost simultaneously. □

Received 24 June; accepted 25 October 1993.

1. Nicolas, A. *Structures in Ophiolites and Dynamics of Oceanic Lithosphere* (Kluwer, Dordrecht, 1989).
2. Nicolas, A., Boudier, F. & Ildefonse, B. in *Magmatic Systems* (ed. Ryan, M. P.) (Academic, Orlando, in the press).
3. Kelemen, P. B., Dick, H. J. B. & Quick, J. E. *Nature* **358,** 635–641 (1992).
4. Nehlig, P. & Juteau, T. *Tectonophysics* **151,** 199–221 (1988).
5. Nicolas, A. & Jackson, M. *J. Petrology* **23,** 568–582 (1982).
6. Sleep, N. H. *J. geophys. Res.* **93,** 10255–10272 (1988).
7. Foulger, G. R. *et al. Nature* **358,** 488–490(1992).
8. Rubin, A. M. *Bull. Volcan.* **52,** 302–319 (1990).
9. Rabinowicz, M., Ceuleneer, G. & Nicolas, A. *J. geophys. Res.* **92,** 3475–3486 (1987).
10. Su, W. & Buck, R. *J. geophys. Res.* **98,** 12191–12205 (1993).
11. Scott, D. R. & Stevenson, D. J. *J. geophys. Res.* **94,** 2973–2988 (1989).
12. Nicolas, A., Freydier, C., Godard, M. & Vauchez, A. *Geology* **21,** 53–56 (1993).

ACKNOWLEDGMENTS. We thank N. Sleep and A. Rubin for reviews of this manuscript, and A. Vauchez and B. Celerier for discussions and comments. We acknowledge contributions to field measurements by E. Gnos, G. Ceuleneer, I. Reuber, M. Misseri, K. Benn and C. MacLeod. The term 'breathing ridge' was suggested by J. Pallister.

Seismic anisotropy in the mantle beneath an oceanic spreading centre

Donna K. Blackman*, John A. Orcutt*, Donald W. Forsyth† & J-Michael Kendall*‡

* Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, La Jolla, California 92093-0225, USA
† Department of Geological Sciences, Brown University, Providence, Rhode Island 02912, USA

BENEATH an active mid-ocean ridge, the mantle upwells in response to the divergence of the newly formed plates, leading to high temperatures and pressure-release melting below the ridge axis. The width of the upwelling region and the amount of melting depend on mantle rheology[1–5], but all models predict a maximum decrease in seismic velocity at the ridge axis. It has also been suggested, however, that the alignment of anisotropic minerals by shear in the upwelling mantle will increase seismic velocity for rays travelling subvertically through the upwelling zone[6,7]. Here we report the observation of a consistent pattern of anomalously early P-wave arrival times at an array of ocean-bottom seismographs deployed across the axis of the southern Mid-Atlantic Ridge: P-waves from distant earthquakes arrive earlier at stations near the axis than at those further away. Our results are consistent with a model of anisotropy in which the degree of mineral alignment is greatest directly beneath the ridge axis, and significant anisotropy extends tens of kilometres from the axis.

Three earthquakes of magnitude greater than 6 were recorded during a 4-week deployment of ocean-bottom seismographs (OBSs) at the Mid-Atlantic Ridge (MAR; 34° S) in Spring 1990. The OBS array consisted of six instruments, three spaced ~25 km apart of the western flank of the ridge and three similarly spaced on the eastern ridge flank (Fig. 1). Relative plate velocities average 35 km Myr^{-1} along this section of the MAR and the morphology is characteristic of a slow-spreading centre[8–10].

The teleseismic data comprise 19 long-period recordings of P-phases from 3 sources (Table 1). Cross-correlation of the individual phases was used to determine the relative travel times to different instruments in the array. In every case, the near-axis OBSs record the P-waves 0.4–1.5 s earlier than the off-axis stations (Table 1 and Fig. 2). Within the data uncertainty of 0.25 s, an increase in relative delay with distance from the axis is evident: 0.6 s delay ~50 km off-axis; 1.2 s ~75 km off-axis.

TABLE 1 Relative travel time of P-phases

Earthquake P-phase*	ΔFRD† (s)	Xcorr‡ (s)	Delay§ (s)
0405 PKP			
OPS	+0.42	+0.50	0.08
FRD	0	0	0
KRN	−0.12	+0.66	0.78
JDY	−0.36	+0.75	1.11
0405 PP			
FRD	0	0	0
KRN	−0.54	+0.06	0.60
JDY	−1.32	+0.06	1.38
0418 PKP			
JAN	+1.08	+0.13	0.95
FRD	0	0	0
KRN	−0.30	+0.13	0.43
JDY	−0.66	+0.16	0.82
0418 PP			
JAN	+3.72	+4.12	0.40
FRD	0	0	0
KRN	−0.96	−0.25	0.71
JDY	−2.10	−0.62	1.48
0403 P			
JAN	−3.42	−2.84	0.58
FRD	0	0	0
KRN	+0.90	+2.34	1.44
JDY	+1.68	+2.59	0.91

Relative travel times of P-phases determined from differential pressure gauge records at seismometers JAN, OPS, FRD, KRN and JDY. (During earthquake 0405, JAN was in the process of writing to tape so did not record the event. OPS had inconsistencies in sampling rate so only the unfiltered data for the largest event (earthquake 0405) was used from this instrument.)

* P-phase and earthquake for which the listed instruments have data.

† Travel time relative to ocean-bottom seismograph (OBS) FRD as predicted by reference Earth model PREM.

‡ Observed travel times relative to FRD as determined from cross-correlation of waveforms.

§ Travel-time anomaly relative to PREM; arrivals are earliest at the near-axis stations FRD and OPS.

The main contribution to the travel-time anomalies across the OBS array comes from the uppermost mantle. Near-source and along-path effects are eliminated as the paths of the rays are essentially identical at teleseismic distances and we use relative travel times. Variations in mantle temperature or the thickness of the crust could affect travel times, but we estimate the size of

‡ Present address: Department of Physics, University of Toronto, Toronto, Ontario M5S 1A7, Canada.

these contributions to be similar in magnitude to the uncertainties in our analysis. Gravity anomalies inferred to reflect crustal thickness variations beneath our OBS array[11] suggest corresponding travel-time anomalies of <0.25 s. Advection of hot mantle from depth results in higher temperatures, therefore slower seismic velocities, in the upwelling region compared to the surrounding mantle[12,13]. Converting oceanic PP-P residuals[14] to one-way vertical travel times versus plate age (-0.07 s $Myr^{-1/2}$) gives an axial thermal delay of only 0.15 s relative to a distance off-axis equal to the half-width of our array.

More important than the solid-state effect of high subaxial temperatures is the partial melting that occurs. The resulting basaltic liquid has significantly lower seismic velocities than ambient mantle, but its detectability depends critically on how the melt is distributed in the host rock[15]. The presence of as little as 1% melt fraction distributed in a triangular region of 120-km wide base at 60 km depth will delay P-waves by ~0.1 s (refs 16, 17). The fact that we do not find slow near-axis arrivals in our data suggests that either the amount of melt present at this location is small or that the melt zone extends <25 km off-axis. The presence of a narrow melt zone may be suggested by a previous measurement of teleseismic P-wave delay at the northern MAR, where arrivals at an instrument within the axial median valley were 0.5-s later than arrivals at the rift walls[18].

Shear stresses that develop during viscous flow of the mantle are predicted to result in seismic anisotropy in the deformed peridotite. Field studies in ophiolites indicate that the *a*-axis of olivine crystals aligns towards the direction of the shear plane during deformation of the mantle[19,20]. Compressional seismic energy travels along the *a*-axis of olivine 22% faster than it does along the *b*-axis[20]. As the dominant shear plane beneath a (two-dimensional) spreading centre is vertical, a detectable travel-time anomaly could accrue as P-waves travel through the upwelling zone. It is not yet clear exactly how factors that could influence subaxial anisotropy interact (for example, mantle stress field, the degree to which olivine crystals within an aggregate align, the dominant direction of that alignment and the change in elastic constants) but results so far indicate that it is a potentially significant contributor to the relative travel times across a spreading centre[17,20–22].

Seismic modelling shows that the observed near-axis travel-time advance can be reasonably attributed to anisotropy in the upwelling region for a plausible spreading-centre velocity structure (Fig. 2). The model shown is one in a series for which the travel-time contributions of melt, lithospheric anisotropy, and upwelling mantle anisotropy were evaluated[17]. Seismic rays are traced through the model with a computer programme that incorporates the effects of three-dimensional heterogeneous, anisotropic structure[23]. Rays are traced initially through a spherically symmetric global model, PREM, for ranges comparable to the observed earthquakes (Fig. 1, top). At a depth of 80 km, rays spaced a few kilometres apart enter the bottom of the prescribed ridge velocity structure. The model that best fits the southern MAR data contains a lithosphere that thickens off-axis at a rate appropriate for the spreading rate of 18 mm yr^{-1} and has 7% anisotropy with the fast-axis horizontal. There is no melt present and the only temperature condition is that embodied in the definition of lithospheric thickness which increases with the square root of plate age. The anisotropy due to upwelling is prescribed in a region centred on the axis that has a maximum value out to 30 km, then is cosine-tapered to <1% (essentially isotropic) at 60 km off-axis; 15% anisotropy in the upwelling region is required to fit the data.

There is some trade-off between the amount of anisotropy determined for the upwelling region, its depth, and the anisotropy in the lithosphere, but we cannot fit the data without significant vertical anisotropy near the axis. The choice of 7% horizontal anisotropy in the lithosphere is based on several previous studies (for example, refs 24–27) that determined the average anisotropy for large areas. It is possible that these average values contain locally greater degrees of lithospheric anisotropy. We could reduce the required amount of upwelling anisotropy by a few per cent if we increased the lithospheric anisotropy, but the smooth nature of the increase in plate thickness off-axis precludes a major reduction in upwelling anisotropy because the data indicate that the travel-time anomaly is relatively narrow. If we doubled the depth of the anisotropic upwelling we could reduce the required degree of anisotropy to a value similar to that in the lithosphere. The depth to which spreading-centre dynamics control the mantle stress field is not well constrained; our choice of 80 km as the base of the model reflects the depth at which significant across-axis variability may begin to develop due to the onset of melting.

The 60-km half-width of the anisotropic upwelling region is constrained to within <10 km for the simple geometry assumed. On a global scale, the width of this inferred region of high anisotropy is quite small. It would be difficult to resolve this feature using land seismic stations because the bounce-point coverage does not typically span spreading axes. This, combined with the use of S-waves which show little anomalous behaviour in response to vertically aligned crystal anisotropy for near-vertical raypaths[17], may be why previous SS-S studies did not detect an axial anomaly[27,28]. Our inferred degree of upwelling anisotropy seems high but this value is not unprecedented. Wide-angle reflection data off Norway can be explained by 15% anisotropy in the upper mantle[29] although in this case the anisotropy is attributed to shearing associated with a dipping master fault

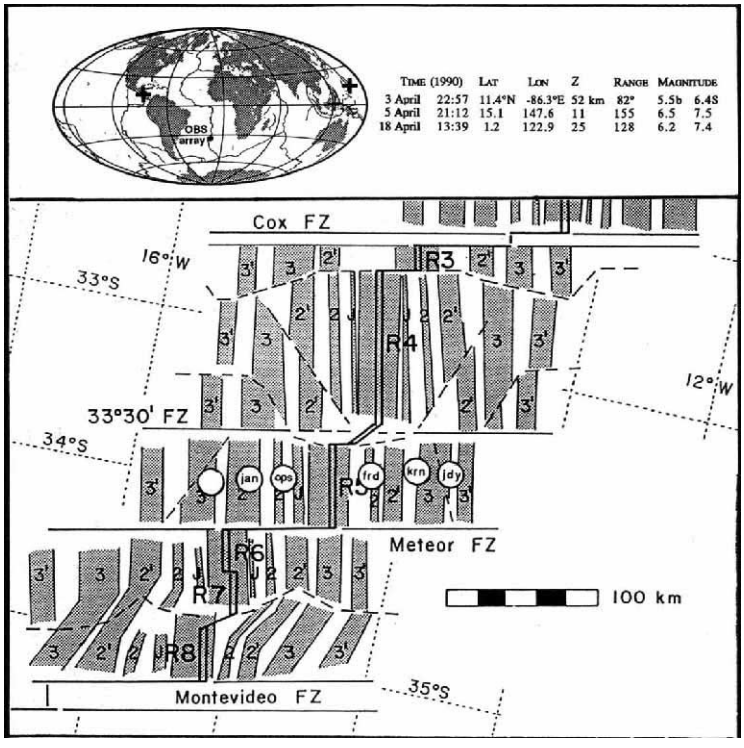


FIG. 1 Location of seismometer array. Top, map showing the position of the array along the mid-ocean ridge and also the locations (+) of the 3 earthquakes used in our analysis. Bottom, diagram showing the tectonic setting of the experiment within a series of ridge segments R (double heavy lines) and fracture zones FZ (light lines). Magnetic anomalies are identified on shaded stripes. Specific instrument names are circled to indicate their position relative to the spreading axis (the instrument on the extreme left is not discussed in this paper because the differential pressure-gauge channel was unreadable).

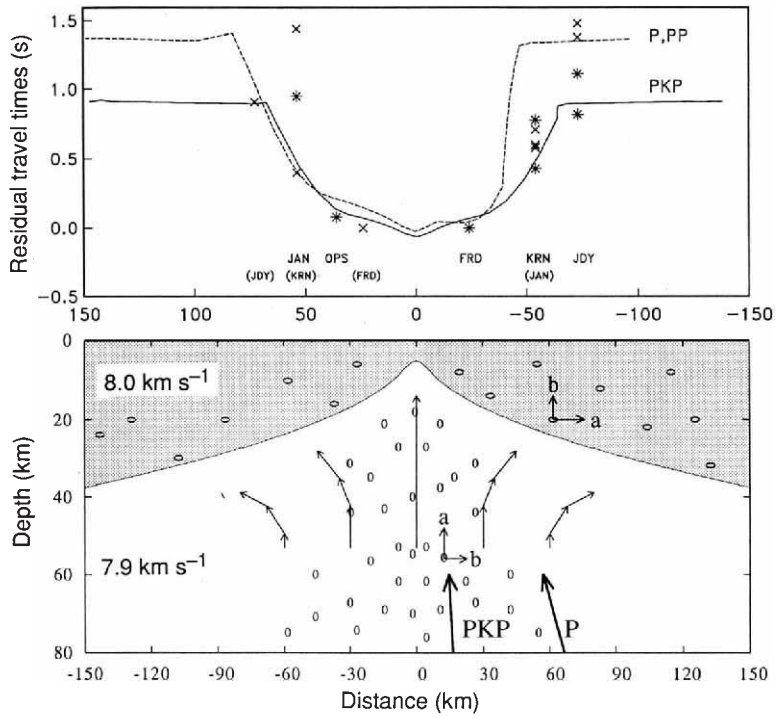


FIG. 2 Comparison of observed and modelled P-wave travel times at a slow-spreading centre. Bottom, sketch illustrating mantle flow (light arrows), off-axis thickening of plates (shaded), and anisotropy in upwelling region (ellipses). The long axis of an ellipse indicates the orientation of the *a*-axis of olivine, which is the fast direction for P-waves; the *b*-axis lies in the plane of the figure and the *c*-axis is perpendicular to this plane. The density of ellipses illustrates the degree of anisotropy which reaches a maximum of 15% in the subaxial region and tapers to <1% between 30 and 60 km off-axis. Anisotropy frozen into the plate (7%) is horizontal due to rotation of stresses near the surface. Angle of P-waves at the base of the model is indicated by heavy arrows. Top, residual travel times calculated for densely-spaced rays are plotted for PKP- (solid line) and PP-, P- (broken line) arrivals. PKP phase data shown by an asterisk, PP and P phases by a cross. Model delays are for phases traversing upwards from right to left so, for consistency, the data from earthquake 0403 are flipped with respect to the axis; the corresponding OBS names are shown in parentheses. The misfit between the 0403 P-wave at KRN and the model may be due to delays associated with travel through the ridge-offset region between R4 and R5 (Fig. 1).

that formed during regional extension. Recent modelling of the response of polycrystalline mantle rock to shear flow[30] suggests that the high degree of olivine alignment that develops during convective upwelling may be diffused as the flow turns to follow the direction of plate spreading. Thus, the estimated 7–8% of lithospheric anisotropy may not reflect the maximum possible mantle values.

Our data are few in number and site-specific, so we do not wish to interpret the results as a general indication that crystal alignment due to shear flow overwhelms other signals at all mid-ocean ridges. At slow-spreading systems, enhancement of the vertical flow rates (and presumably a corresponding increase in degree of anisotropy) in a narrow region beneath the axis can be caused by buoyancy-driven upwelling[4,5]. Conversely, at a fast-spreading centre, the high upwelling rates tend to overwhelm buoyancy-driven flow, so we might expect smaller relative travel-time anomalies due to shear-induced anisotropy than are observed at the southern MAR.

At present we are not able to use the shape of the observed travel-time anomaly to tightly constrain the width of the upwelling zone, because of uncertainties in the relationship between mantle shear, aligned melt and the development of seismic anisotropy. If the qualitative model of seismic anisotropy oriented in the direction of flow is correct, then the 60–70 km wide early-arrival region may reflect enhanced upwelling due to the presence of high melt fractions (for example, ref. 1). On the other hand, initial quantitative models of subaxial mantle structure and seismic anisotropy suggest that deformation can result in fast-axis orientations that are inclined to the flow direction by 30°–40° (refs 21, 22, 30) and that the presence of aligned melt films or melt-filled cracks can produce anisotropic behaviour as well[17]. A combination of additional teleseismic observations at mid-ocean ridges and theoretical modelling of the possible waveform effects of anisotropy distributions produced by different mechanisms are required in order to fully interpret the travel-time signal at the sea floor. The most diagnostic experiments will record both P- and S-phases on OBS arrays that span the axis with receivers both in the axial region and extending well beyond the half-width predicted for the upwelling zone by passive flow models at the appropriate spreading rate. □

Received 9 August; accepted 22 October 1993.

1. Buck, W. R. & Su, W. *Geophys. Res. Lett.* **16,** 641–644 (1989).
2. Sotin, C. & Parmentier, E. M. *Geophys. Res. Lett.* **16,** 835–838 (1989).
3. Scott, D. R. & Stevenson, D. J. *J. geophys. Res.* **94,** 2973–2988 (1989).
4. Turcotte, D. L. & Phipps Morgan, J. in *Mantle Flow and Melt Generation at Mid-Ocean Ridges* (eds Phipps Morgan, J., Blackman, D. K. & Sinton, J. M.) 155–182 (Am. geophys. Un., Washington DC, 1992).
5. Scott, D. R. in *Mantle Flow and Melt Generation at Mid-Ocean Ridges* (eds Phipps Morgan, J., Blackman, D. K. & Sinton, J. M.) 327–352 (Am. geophys. Un., Washington DC, 1992).
6. Anderson, D. L. *Theory of the Earth* (Blackwell Scientific, Boston, 1989).
7. Forsyth, D. W. in *Mantle Flow and Melt Generation at Mid-Ocean Ridges* (eds Phipps Morgan, J., Blackman, D. K. & Sinton, J. M.) 1–66 (Am. geophys. Un., Washington DC, 1992).
8. Macdonald, K. C. *A. Rev. Earth planet. Sci.* **10,** 155–190 (1982).
9. Kuo, B-Y., Forsyth, D. W. *Mar. Geophys. Res.* **10,** 205–232 (1988).
10. Carbotte, S., Welch, S. M. & Macdonald, K. C. *Mar. Geophys. Res.* **13,** 51–80 (1991).
11. Neumann, G. A. & Forsyth, D. W. *J. geophys. Res.* **98,** 17891–17910 (1993).
12. Sleep, N. *J. geophys. Res.* **74,** 542–549 (1969).
13. Phipps Morgan, J. & Forsyth, D. W. *J. geophys. Res.* **93,** 2955–2966 (1988).
14. Woodward, R. L. & Masters, G. *J. geophys. Res.* **96,** 6351–6377 (1991).
15. Schmeling, H. *Phys. Earth planet. Inter.* **41,** 34–57 (1985).
16. Faul, U., Toomey, D. R. & Humphreys, E. *Seismic Imaging of the East Pacific Rise Upper Mantle* (RIDGE Office Report, Woods Hole, 1992).
17. Kendall, J-M. *Geophys. Res. Lett.* (in the press).
18. Rowlett, H. & Forsyth, D. W. *Geophys. Res. Lett.* **6,** 273–276 (1979).
19. Nicolas, A., Boudier, F. & Boullier, A. M. *Am. J. Sci.* **273,** 853–876 (1973).
20. Christensen, N. I. *Geophys. J. R. Astr. Soc.* **76,** 89–111 (1984).
21. Ribe, N. M. *J. geophys. Res.* **94,** 4123–4223 (1989).
22. Wenk, H-R., Bennett, K., Canova, G. R. & Molinari, A. *J. geophys. Res.* **96,** 8337–8349 (1991).
23. Guest, W. S. & Kendall, J-M. *Can. J. Explor. Geophys.* (in the press).
24. Raitt, R. W., Shor, G. G., Francis, T. J. G. & Morris, G. B. *J. geophys. Res.* **74,** 3095–3109 (1969).
25. Forsyth, D. W. *Geophys. J. R. astr. Soc.* **43,** 103–162 (1975).
26. Nishimura, C. E. & Forsyth, D. W. *Geophys. J. Int.* **94,** 479–501 (1988).
27. Kuo, B-Y., Forsyth, D. W. & Wysession, M. *J. geophys. Res.* **92,** 6421–6436 (1987).
28. Sheehan, A. F. & Solomon, S. C. *J. geophys. Res.* **96,** 19981–20009 (1991).
29. Mjelde, R. & Sellevoll, M. A. *Tectonophysics* **222,** 21–32 (1993).
30. Chastel, T. B., Dawson, P. R., Wenk, H-R. & Bennett, K. *J. geophys. Res.* (in the press).

Support for anisotropy of the Earth's inner core from free oscillations

Jeroen Tromp

Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts 02138, USA

IN 1983, Poupinet *et al.*[1] observed that compressional seismic waves traversing the inner core along a trajectory parallel to the Earth's rotation axis arrive faster than the same (PKIKP) waves travelling in the equatorial plane. They interpreted this observation as revealing prolate topography of the inner-core boundary. In 1986, Morelli *et al.*[2] and Woodhouse *et al.*[3] suggested that inner-core anisotropy could explain both the travel-time observations and the anomalous splitting of some of the Earth's normal modes. Inner-core anisotropy continues to be the preferred explanation for the travel-time anomalies, although there is disagreement about the magnitude of anisotropy[4–6]. More recent explanations for the anomalous splitting involve topography of the inner-core and core–mantle boundaries as well as lateral heterogeneity of the core[7–11]. In particular, Widmer *et al.*[11] dismissed a rather complex recent model of inner-core anisotropy[12] because it could not explain the splitting of several previously unidentified modes. Here I show that the anomalous splitting of all currently identified modes can in fact be explained by cylindrical anisotropy of the Earth's inner core that is also compatible with the observed PKIKP travel-time anomalies. The resulting model should be regarded as an upper limit to the amount of anisotropy, as lateral heterogeneity also undoubtedly contributes to the splitting.

The evidence for inner-core anisotropy from travel-time data is accumulating but remains inconclusive. Shearer and coworkers[4,5] proposed uniform anisotropy at the 1% level throughout the inner core, because their PKIKP travel-time observations at distances <155° were smaller than the predictions of the Morelli *et al.*[2] model. More recently, Creager[6] showed that there are significant travel-time anomalies between rays turning in the liquid outer core (PKP-BC) and rays turning in the solid inner core (PKIKP) near 150°, contradicting the findings of Shearer *et al.*[4,5] The model of Creager[6] requires some 3.5% anisotropy near the top of the inner core, but his data control only the outermost 300 km, and the anisotropy must vanish rapidly below that depth for his model to yield predictions consistent with observations at 180°. The results of Creager[6] are conceptually compatible with those in Morelli *et al.*[2] and Woodhouse *et al.*[3] The main difference between the analysis of Shearer and coworkers[4,5] and Creager[6] is that the former authors used data from the International Seismological Center (ISC), whereas Creager hand-picked PKP-BC minus PKIKP travel-time anomalies. Support for inner-core anisotropy from normal-mode observations appears to be even less conclusive, judging from the incompatible results of Woodhouse *et al.*[3], Li *et al.*[12] and Widmer *et al.*[11]. Because of these discordant claims and results the concept of inner-core anisotropy is in doubt. My aim here is to determine if inner-core anisotropy compatible with the travel-time observations is capable of explaining the splitting observations, or whether another—as yet unknown—cause is required.

A free oscillation or normal mode of the Earth can be characterized by an eigenfunction and an associated eigenfrequency ${}_n\omega_l^m$. The overtone number n angular degree l and azimuthal order m are integers which are used to identify a specific mode of oscillation. For every value of l there are $2l+1$ associated values of m: $m=-l,\ldots,m=0,\ldots,m=l$. A multiplet ${}_nS_l$ (spheroidal modes) or ${}_nT_l$ (toroidal modes) is the collection of all $2l+1$ free oscillations with the same quantum numbers n and l; the $2l+1$ members of a multiplet are called singlets. On a spherically symmetric Earth model such as PREM[13], all singlets within a given multiplet have the same eigenfrequency ${}_n\omega_l$; we say that the singlets are $2l+1$ degenerate. Any departure of the Earth from sphericity removes this degeneracy and causes the singlets to split, such that each individual singlet has its own distinct eigenfrequency ${}_n\omega_l^m$.

The Earth's largest deviations from sphericity are its rotation and ellipticity of figure. The associated splitting is predicted to be of the form[14]

$$\omega^m = \omega_c(1 + bm + cm^2) \qquad (1)$$

where the indices n and l have been omitted for clarity. The parameter ω_c is referred to as the centre frequency of the multi-

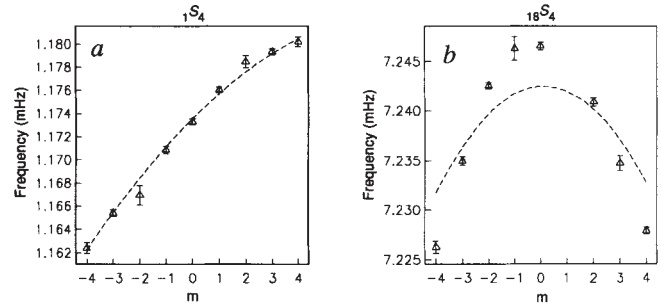


FIG. 1 *a*, Observed and predicted splitting for spheroidal mode ${}_1S_4$. *b*, Observed and predicted splitting for the anomalously split spheroidal mode ${}_{18}S_4$. Observed singlet eigenfrequencies are denoted by triangles. Predicted splitting as a result of the Earth's rotation and ellipticity of figure, equation (1), is indicated by the dashed line.

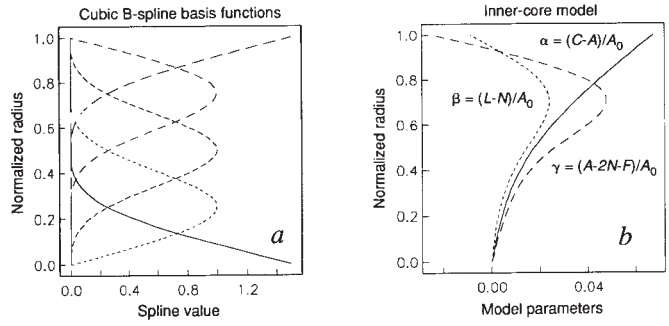


FIG. 2 *a*, Cubic B-spline basis functions used in the inversion[18]. A normalized radius of 1 corresponds to the inner-core boundary. The advantage of this representation is that it allows for models that are relatively concentrated in particular regions of the inner core. We are unable to produce a more detailed model of the inner core because of the relatively small data set and because the free oscillations that constitute the data set have only limited radial resolution. The spline functions are used to construct the anisotropic inner-core model. *b*, Anisotropic inner-core model obtained by inversion of the anomalously split modes identified by Ritzwoller *et al.*[7,9], Giardini *et al.*[10] and Widmer *et al.*[12]. A normalized radius of 1 corresponds to the inner-core boundary. The parameter α shows an overall increase with radius to a maximum value of 6.8% at the inner-core boundary. The parameters β and γ show an overall decrease with radius to minimum values of −0.9 and −2.3%, respectively. The overall behaviour of the parameters α, β and γ with depth is similar to the behaviour of the same parameters in an inner-core model presented by Woodhouse, Giardini and Li[3]. In that model, based on the seven anomalously split modes ${}_3S_2$, ${}_6S_3$, ${}_9S_3$, ${}_{11}S_4$, ${}_{11}S_5$, ${}_{13}S_2$ and ${}_{13}S_3$, the parameters α, β and γ take on the values 10.4, 1.9 and −3.3%, respectively, at the inner-core boundary and have a prescribed quadratic radial dependence. For a given inner-core model, the centre frequency ω_c is determined separately for each individual multiplet by least-square fitting the observed singlet eigenfrequencies to the predicted singlet eigen frequencies. The 15 spline coefficients are determined by the downhill simplex method[21] which does not require the calculation of Fréchét derivatives.

plet. The first-order effects of the Earth's rotation are represented by the parameter b and produce linear splitting as a function of m. The Earth's ellipticity of figure and the second-order effects of its rotation are represented by the parameter c and produce quadratic splitting in m. Figure 1a shows an example of the observed and predicted splitting of spheroidal mode ${}_1S_4$ which is predominantly split by rotation.

Masters and Gilbert[15] were the first to identify a collection of modes that is split much more than predicted from the Earth's rotation and ellipticity of figure. Figure 1b shows an example of an anomalously split mode; this enhanced splitting is characteristic of all such free oscillations. Over the last decade the number of observed anomalously split modes has grown to a total of ~20 (refs 3, 7–12). There seems to be general agreement that the structure responsible for the observed splitting is cylindrically symmetric about the Earth's rotation axis and located at or below the core–mantle boundary. Three mechanisms to explain the anomalous splitting have been proposed. The first[7-11]

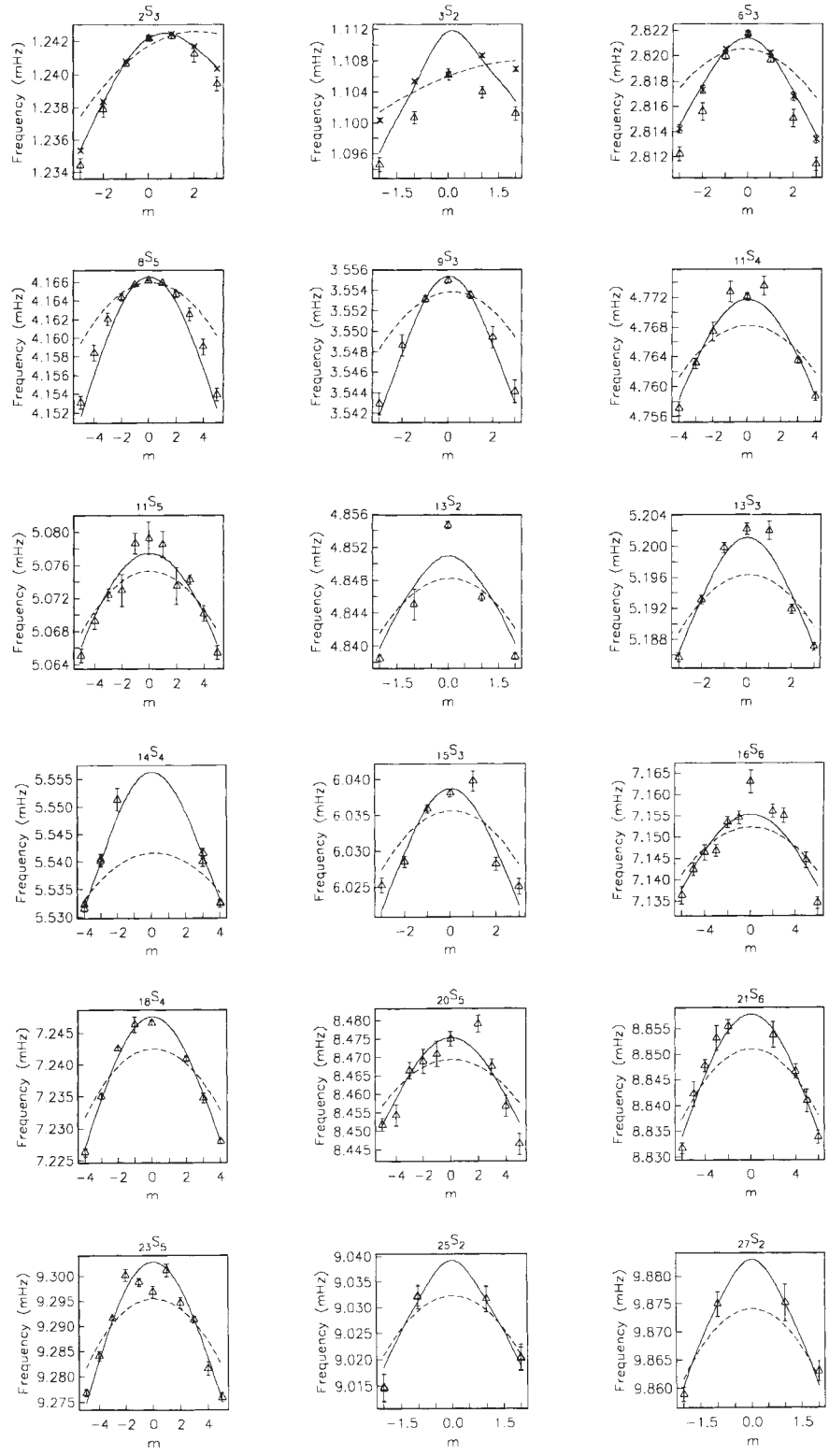


FIG. 3 Observed and predicted splitting for the data set consisting of 18 anomalously split free oscillations. For modes ${}_2S_3$, ${}_3S_2$, ${}_6S_3$, ${}_8S_5$ and ${}_9S_3$, triangles indicate singlet eigenfrequencies obtained from the splitting coefficients published by Ritzwoller *et al.*[7,9]. For modes ${}_2S_3$, ${}_3S_2$, and ${}_6S_3$, crosses indicate singlet eigenfrequencies obtained from the splitting coefficients published by Giardini *et al.*[10]. Finally, for modes ${}_{11}S_4$, ${}_{11}S_5$, ${}_{13}S_2$, ${}_{13}S_3$, ${}_{14}S_4$, ${}_{15}S_3$, ${}_{16}S_6$, ${}_{18}S_4$, ${}_{20}S_5$, ${}_{21}S_6$, ${}_{23}S_5$, ${}_{25}S_2$, and ${}_{27}S_2$, triangles indicate singlet eigenfrequencies determined by Widmer *et al.*[12]. Predicted splitting as a result of the Earth's rotation and ellipticity of figure, equation (1), is indicated by the dashed line, and theoretical splitting as a result of rotation, ellipticity, and cylindrical anisotropy of the Earth's inner core, equation (2), is indicated by the solid line. Notice the discrepancy in observed singlet eigenfrequencies for mode ${}_3S_2$, which is one of the few modes that is sensitive to the shear-speed structure of the inner core.

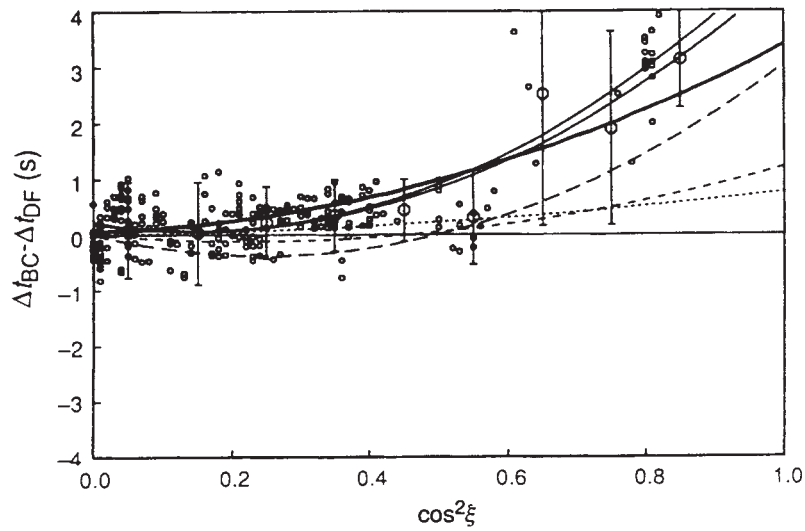


FIG. 4 Predicted and observed PKP-BC minus PKIKP-DF travel-time anomalies due to the anisotropy shown in Fig. 2*b* plotted as a function of $\cos^2\xi$, where ξ is the angle between the inner-core leg of the PKIKP ray and the Earth's rotation axis. (This Figure is based upon Fig. 2*a* of Creager[6].) The observed anomalies, indicated by the small circles, were determined by Creager[6]. Large circles and bars show mean and two standard deviations of the data binned in 1° intervals. Travel-time anomalies are calculated based upon equation (3) for surface-focus events at an epicentral distance of ~150°. Thick solid line, this study. The following curves were calculated by Creager[6]: long-dashed line, ref. 2; short-dashed line, ref. 4; dotted line, ref. 5; two thin solid lines, ref. 6.

involves lateral heterogeneity of the core proportional to the spherical harmonic Y_2^0. Although such models seem to reproduce most of the observed splitting, these studies recognize that the inferred lateral heterogeneity of the core is physically unrealistic[16]. The second mechanism involves topography on the inner core and core-mantle boundaries[1,7,9,10]. With the observed PKIKP travel-time anomalies in mind, the third proposed mechanism[3,12] involves anisotropy of the inner core that exhibits cylindrical symmetry about the Earth's rotation axis. Topography and lateral heterogeneity seem to be incapable of realistically explaining the PKIKP travel-time observations.

Here I reinvestigate the possibility of a cylindrically anisotropic inner core from the point of view of free oscillations based upon the currently identified anomalously split modes. The simplest type of anistropy that exhibits cylindrical symmetry about the Earth's rotation axis is transverse isotropy, which involves the five elastic parameters A, C, F, L and N (ref. 17). The parameters C and A correspond to the speed of compressional waves travelling parallel or perpendicular to the Earth's rotation axis, respectively, whereas the parameters L and N correspond to the speed of shear waves travelling parallel or perpendicular to the same axis. The parameter F influences the speed of waves travelling at other angles with the rotation axis. It can be demonstrated (J.T., manuscript in preparation) that this type of anisotropy together with rotation and ellipticity produces splitting of the form

$$\omega^m = \omega_c'(1 + bm + cm^2 + c'm^2 + dm^4) \qquad (2)$$

The coefficients c' and d represent the effects of inner-core anisotropy and are completely determined by the three parameters $\alpha = (C-A)/A_0$, $\beta = (L-N)/A_0$, and $\gamma = (A-2N-F)/A_0$, where $A_0 = \kappa + \frac{4}{3}\mu$ is determined by the bulk modulus κ and shear modulus μ at the centre of the reference Earth model. When $\alpha > 0$, inner-core P waves travel slower in the equatorial plane than along the Earth's rotation axis. Similarly, when $\beta > 0$, inner-core S waves travel slower in the equatorial plane than along the Earth's rotation axis; such inner-core S waves have never been unambiguously observed. The third parameter, γ, influences the speed of P and S waves travelling in directions other than parallel or perpendicular to the axis of rotation. Using a data set collected by Ritzwoller *et al.*[7,9], Giardini *et al.*[10] and Widmer *et al.*[11] consisting of the 18 anomalously split multiplets ${}_2S_3$, ${}_3S_2$, ${}_6S_3$, ${}_8S_5$, ${}_9S_3$, ${}_{11}S_4$, ${}_{11}S_5$, ${}_{13}S_2$, ${}_{13}S_3$, ${}_{14}S_4$, ${}_{15}S_3$, ${}_{16}S_6$, ${}_{18}S_4$, ${}_{20}S_5$, ${}_{21}S_6$, ${}_{23}S_5$, ${}_{25}S_5$, ${}_{25}S_2$ and ${}_{27}S_2$, for a total of 145 singlet eigenfrequencies, I invert for the radial dependence of α, β and γ.

I represent the radial dependence of the three model parameters α, β and γ in terms of five cubic B-spline basis functions[18], as shown in Fig. 2*a*. I seek to determine five spline coefficients for each of the three model parameters, which makes for a total of 15 unknowns. Figure 2*b* shows the resulting inner-core model. The splitting predicted for the 18 anomalously split modes in the data set is shown in Fig. 3. The reduction in variance achieved by the model in Fig. 2*b* compared to a model incorporating only the effects of rotation and ellipticity is 72%. The value of χ^2/ν, where $\nu = 145 - 15 = 130$ is the number of degrees of freedom, is 17.0 for the initial model based upon rotation and ellipticity and 3.4 for the final model based upon rotation, ellipticity and inner-core anisotropy. The details of the top of the inner-core model in Fig. 2*b* were based on recent travel-time observations (W.-J. Su and A. Dziewonski, manuscript in preparation), particularly in the epicentral distance range 130–136°. It is interesting to note that modes such as ${}_{11}S_4$, ${}_{11}S_5$ and ${}_{16}S_6$, which have <2% of their energy in the inner core, are nevertheless reasonably well fitted by the model. (Widmer *et al.*[11] used the relatively small percentage of inner-core energy of some anomalously split modes as an argument against inner-core anisotropy.) Note, also, that the inner-core model of Woodhouse *et al.*[3], which is compatible with the model in Fig. 2*b*, produces the observed splitting with a variance reduction of 56%. An anisotropic inner-core model proposed by Creager[6], which was designed to fit travel-time anomalies, can also produce the observed splitting, with a similar variance reduction, provided I restrict Creager's uniform anisotropy to the upper 300 km of the inner core.

The P velocity (v) perturbation associated with the model shown in Fig. 2*b* is given by[2]

$$\frac{\delta v}{v} = (2\beta - \gamma)\cos^2\xi + (\tfrac{1}{2}\alpha - 2\beta + \gamma)(\cos^2\xi)^2 \qquad (3)$$

where ξ is the angle between the PKIKP ray segment in the inner core, and the Earth's rotation axis. Although travel-time observations are not included in the inversion, I do force it to look for models that are compatible with those observations; this procedure does not significantly affect the fit to the normal-mode data. In Fig. 4 the travel-time anomaly between rays turning in the outer core (PKP-BC) and those turning in the inner core (PKIKP) due to the anisotropy shown in Fig. 2*b* is plotted as a function of $\cos^2\xi$. The predicted travel-time anomalies agree quite well with the observed anomalies of Creager[6]. PKIKP travel times of waves travelling parallel to the Earth's rotation axis are 2.6 s faster than for waves travelling in the equatorial plane, which is also on the order of the observed PKIKP travelling time anomalies[2].

To decide the issue of inner-core anisotropy based upon travel times probably requires taking a new look at the records used

by the ISC. If we are willing to accept the hypothesis that the inner core exhibits cylindrical anisotropy, this raises the question of what physical process could be responsible for this phenomenon. Previous studies[19,20] suggest that the inner core consists of iron in its hexagonally close-packed phase; this ε-phase of iron has cylindrical symmetry, which would be in agreement with the results described here. □

Received 20 May; accepted 19 October 1993.

1. Poupinet, G., Pillet, R. & Souriau, A. *Nature* **305,** 204–206 (1983).
2. Morelli, A., Dziewonski, A. M. & Woodhouse, J. H. *Geophys. Res. Lett.* **13,** 1545–1548 (1986).
3. Woodhouse, J. H., Giardini, D. & Li, X.-D. *Geophys. Res. Lett.* **13,** 1549–1552 (1986).
4. Shearer, P. M. & Toy, K. M. & Orcutt, J. A. *Nature* **333,** 228–232 (1988).
5. Shearer, P. M. & Toy, K. M. *J. geophys. Res.* **96,** 2233–2247 (1991).
6. Creager, K. C. *Nature* **356,** 309–314 (1992).
7. Ritzwoller, M., Masters, G. & Gilbert, F. *J. geophys. Res.* **91,** 10203–10228 (1986).
8. Giardini, D., Li, X.-D. & Woodhouse, J. H. *Nature* **325,** 405–411 (1987).
9. Ritzwoller, M., Masters, G. & Gilbert, F. *J. geophys. Res.* **93,** 6369–6396 (1988).
10. Giardini, D., Li, X.-D. & Woodhouse, J. H. *J. geophys. Res.* **93,** 13716–13742 (1988).
11. Widmer, R., Masters, G. & Gilbert, F. *Geophys. J. Int.* **111,** 559–576 (1992).
12. Li, X.-D., Giardini, D. & Woodhouse, J. H. *J. geophys. Res.* **96,** 551–557 (1991).
13. Dziewonski, A. M. & Anderson, D. L. *Phys. Earth planet. Inter.* **25,** 297–356 (1981).
14. Woodhouse, J. H. & Dahlen, F. A. *Geophys. J. R. astr. Soc.* **53,** 335–354 (1978).
15. Masters, G. & Gilbert, F. *Geophys. Res. Lett.* **8,** 569–571 (1981).
16. Stevenson, D. *Geophys. J. R. astr. Soc.* **88,** 311–319 (1987).
17. Love, A. E. H. *A Treatise on the Mathematical Theory of Elasticity* 4th edn (Cambridge Univ. Press, 1927).
18. Michelini, A. & McEvilly, T. V. *Bull. seism. Soc. Am.* **81,** 524–552 (1991).
19. Brown, J. M. & McQueen, R. G. *J. geophys. Res.* **91,** 7485–7494 (1986).
20. Anderson, O. L. Geophys. *J. R. astr. Soc.* **84,** 561–579 (1986).
21. Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. *Numerical Recipes in C* (Cambridge Univ. Press, 1988).

ACKNOWLEDGEMENTS. I thank R. Widmer for his data set of observed singlet eigenfrequencies and for suggestions regarding those data; K. Creager for his data set of observed travel-time anomalies; W.-J. Su and A. Dziewonski for the opportunity to fit their new data set of PKP-BC minus PKIKP travel-time anomalies; G. Ekström for software for the calculation of cubic B-splines; and A. Dziewonski, G. Ekström, J. Woodhouse, M. Ritzwoller, G. Masters and W.-J. Su for discussions and suggestions. This research was supported by the US National Science Foundation.

Rapid human-induced evolution of insect–host associations

Michael C. Singer*, Chris D. Thomas† & Camille Parmesan*

* Department of Zoology, University of Texas, Austin, Texas 78712, USA
† School of Biological Sciences, University of Birmingham, Edgbaston B15 2TT, UK

RAPID evolution of host association is now occurring independently in two populations of the host-specialist butterfly *Euphydryas editha*, each of which has recently incorporated a novel host species into its diet. The reasons for these episodes of rapid evolution lie in human land use practices: logging in one case and cattle ranching in the other. In contrast to other insects that have used tolerance of human activities to expand their ranges into disturbed habitats[1–3], these rare butterflies have remained at their original sites and evolved adaptations to the changes occurring at those sites. At both sites, the proportion of insects preferring the novel host has increased, in one case clearly because of genetic changes in the insect population. This process is now starting to generate insects that refuse to accept their ancestral host, foreshadowing a new problem in conservation biology. By adapting genetically to human-induced changes in their habitat, the insects risk becoming dependent on continuation of the same practices. This is a serious risk, because human cultural evolution can be even faster than the rapid genetic adaptation that the insects can evidently achieve.

Our study insect, *Euphydryas editha*, belongs to a group of Nymphalid butterflies known in the UK as fritillaries (related to the marsh fritillary) and in the USA as checkerspots. One of the few (<6) known populations of *E. editha monoensis*, which is currently on the list of candidates for endangered status, lives at our Schneider study site (Ormsby Co., Nevada). Evidence for recent diet change by the insects at this site is their use of the European weed, *Plantago lanceolata*, introduced by cattle ranchers[4]. Comparison with nearby *E. editha* populations where *P. lanceolata* has not been introduced suggests that the ancestral diet of Schneider butterflies was *Collinsia parviflora*[4]. Although *C. parviflora* has recently supported much lower larval survival than *P. lanceolata*[5], the Schneider butterflies still use both host species. A third host, *Penstemon rydbergii*, normally receives few eggs.

In nearby sites where *P. lanceolata* has not been introduced, we found no butterflies that preferred it, but about 10% of them accepted it as readily as *C. parviflora*[4]. At Schneider, we did find insects that actively preferred the introduced host, *P. lanceolata*[4].

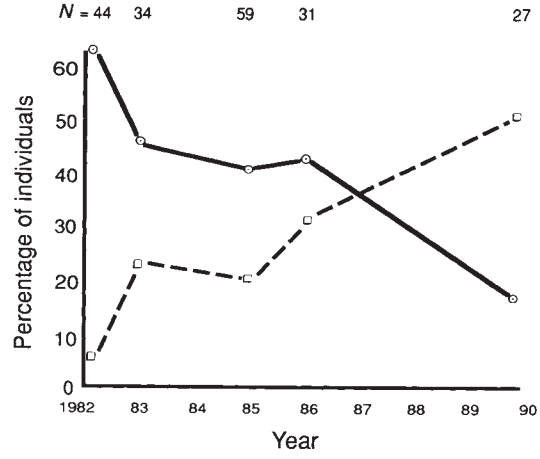


FIG. 1 Changes over time in results of preference trials done in the field at Schneider's Meadow. Each trial was conducted with contemporary plants, undisturbed and still growing in their natural positions. Each insect was repeatedly offered two hosts in alternation, but not allowed to oviposit. Such insects usually pass through three phases: a rejection phase when neither host is accepted, a discrimination phase when one host is consistently accepted and the other consistently rejected, and a final phase when both are accepted until oviposition is allowed by the experimenter[23]. The preferred host is the one that is accepted during the discrimination phase. Data are given for proportions of insects preferring each of the two hosts. The proportion of insects without preference is not presented explicitly, but can be deduced from the figure, because the three categories sum to 1. The ratio of the number of insects preferring *P. lanceolata* (□) to the number preferring *C. parviflora* (⊙) is significantly heterogeneous among years. (Contingency table analysis, $G=27.7$, d.f. $=3$, $P=0.0001$.)

The proportion of these insects in the population has risen sharply in the past decade (Fig. 1), as has the proportion actually using *P. lanceolata* (Table 1). To determine whether genetic change in insect preference had occurred, we compared the preferences of greenhouse-reared offspring of butterflies captured in 1983 and 1990. These preferences differed in the same direction as the change measured in the field (Table 2). This result provides strong evidence for genetic change of preference in the insect population over the seven generations from 1983 to 1990. This is a direct demonstration of the evolution of an insect–plant association. Previously, such changes in non-pest systems were inferred from contemporary patterns, not observed directly[6–16]. One other natural episode of rapid diet-associated evolution has been observed: the bill width of a seed-eating bird, *Geospiza fortis*, decreased following a change in the available diet[17].

Several factors predispose Schneider *E. editha* to rapid diet

evolution. *E. editha* is known for its sedentary nature; this should help the evolution of local adaptation on a relatively fine scale[18]. The Schneider butterflies are isolated by several kilometres from conspecific populations[4]. Selection on diet has been strong[5], heritability of oviposition preference has been high[19], and a population bottleneck occurred during our study (Table 1). In these circumstances, both selection and drift should be effective agents of evolutionary change.

At our second study population, Rabbit (Tulare Co., California), *E. editha* has colonized *Collinsia torreyi*, which was first rendered suitable for the insects by logging about 25 years ago[16]. Clear-cut patches in which this host was used were interspersed with unlogged patches in which the insects still laid eggs on *Pedicularis semibarbata*, their long-standing host. Over the period 1984–86, insect survival from egg to adult was higher on *C. torreyi* in a 'clear-cut' than on *P. semibarbata* in an adjacent unlogged patch[20]. Trials using butterflies emerging in the clear-cut showed an increase in preference for *C. torreyi* and a decrease in preference for *P. semibarbata* from 1979 to 1989 (Fig. 2).

Are the insects at Rabbit and Schneider evolving dependence on human activities? Experiments with nematodes show that dependence on a host that is currently being used can evolve rapidly even without direct selection against use of other hosts[21]. In the case of *Euphydryas*, a complete shift to preference for a novel host need not produce dependence, provided that individuals retain the flexibility to recolonize their long-standing host if they cannot find the novel species. We can measure this flexibility in terms of the time that each insect accepts only its preferred host before reaching the level of motivation at which either host would be accepted, if encountered. Estimates of this length of time, termed a discrimination phase, are obtained in our preference tests (legend to Fig. 1). The longer this discrimination phase, the more likely the insect is to emigrate[22] or die before oviposition occurs, and the more dependent it is on the presence of its preferred host species. Monophagous populations of *E. editha* typically comprise insects with mean discrimination phases of more than 5 hours[23].

All of the Rabbit butterflies we tested, including those that preferred *C. torreyi*, did accept *P. semibarbata*. Mean length of discrimination phase among *C. torreyi*-preferring insects was less than 30 min[24]. Hence, this population is in no immediate danger of becoming dependent on logging by evolving host specificity for *C. torreyi* and losing the ability to lay eggs on *P. semibarbata*.

At Schneider, the first *P. lanceolata*-preferring butterfly with a discrimination phase over 5 hours was found in 1986 (ref. 23). In 1990 we found three very host-specific individuals that declined to accept their ancestral host, even after several days without oviposition. If these insects increase in frequency, the population could become dependent on *P. lanceolata*, an introduced plant which, in the vicinity of Schneider, either disappears or becomes inaccessible to butterflies (overgrown with other vegetation) when humans abandon previously managed land.

P. lanceolata now grows widely in North America, and has also been colonized by *Euphydryas phaeton* in northeastern

TABLE 1 Numbers of egg clusters or larval webs found on *Collinsia*, *Plantago* and *Penstemon* in censuses at Schneider between 1982 and 1993

Year	Area censused (m^2)	Host genus *Collinsia*	*Plantago*	*Penstemon*	Estimated density in clusters per 10,000 m^2
1982	70	28	4	2	4,850
1983	20	11	3	0	7,000
1984	1,250	27	4	0	248
1986	Not recorded	50	62	10	—
1987	2,030	10	8	0	87
1988	20,000	0	3	1	2
1989	50,000	0	23	1	5
1990	2,000	6	25	2	165
1993	5,800	7	22	1	52

The table shows numbers of egg clusters or larval webs (each resulting from a single cluster) found on each host species. Data gathered in different years are not strictly comparable, but we are confident of the population bottleneck in 1988–89 and of the absence of eggs on *C. parviflora* in those years. Between 1987 and 1988 all the insects died except those on a small patch that contained no *C. parviflora*. In 1989 the area occupied was slightly greater, but still did not include patches with *C. parviflora*. In that year we censused the entire area that the population had formerly occupied to confirm that no insects were using *C. parviflora*. Use of this host resumed in 1990, when the population expanded into areas where the plant was present. The ratio of the number of groups found on *P. lanceolata* to the number on *C. parviflora* is significantly heterogeneous among years. (Contingency table analysis, $\chi^2=86.0$, d.f.=8, $P<0.0001$.)

TABLE 2 Numbers of families with mean preferences for either *Collinsia* or *Plantago* among offspring of Schneider insects captured in 1983 and 1990

	Preference for: *Collinsia*	*Plantago*
Family means, offspring of 1983	20	4
Family means, offspring of 1990	3	7

In a previous experiment, we had preference-tested freshly captured females in the field, raised their offspring on *C. parviflora* in a greenhouse, and estimated heritability of preference from a regression of mean daughter's preference on maternal preference for each family[19]. Oviposition preference was significantly heritable in 1983. Family mean preferences were thus available for offspring of insects captured in the field in 1983. The top line of the table telescopes into two categories the data presented as a continuous variable in ref. 19. The mean preference is obtained by averaging the discrimination phases of individuals. Although this does not necessarily follow, the same result would have been obtained in this case by classifying families in terms of the direction of preference shown by the majority of sibs. Here we capitalized on the fact that we still had 8 of the *C. parviflora* clones and 10 first-generation (sexual) offspring of the *P. lanceolata* that had been used in the previous experiment. We used these plants in rotation to apply the same preference tests to offspring of 1990 butterflies that had been raised in the same greenhouse, also on *C. parviflora*.

$P=0.014$ by Fisher's exact test, analysed as a contingency table.

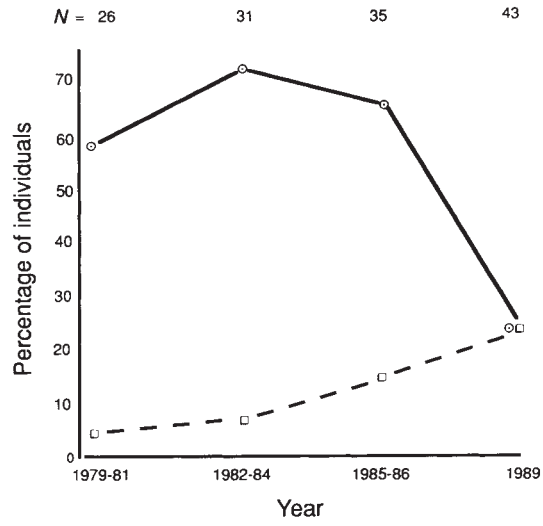


FIG. 2 Changes over time in results of preference trials done in the field at Rabbit Meadow. Each trial was conducted with contemporary plants, undisturbed and still growing in their natural positions. Because larvae do not move far, and because *P. semibarbata* was not present in clear-cut patches, teneral (newly emerged) female butterflies found in clear-cuts can be identified as insects that have developed on *C. torreyi* from eggs naturally laid on that host. By using only insects that had developed on *C. torreyi*, these experiments controlled for effects of larval host species. The ratio of the number of insects preferring *P. semibarbata* (⊙) to the number preferring *C. torreyi* (□) is significantly heterogeneous among years. (Contingency table analysis, $G=16.5$, d.f.=3, $P=0.0007$.)

USA[15]. In California and Nevada, *P. lanceolata* is widespread in disturbed habitats, which the poorly dispersing[18,25] *E. editha* has generally been unable to colonize. Unless the insect evolves to be more dispersive, these habitats seem likely to remain unexploited.

E. editha is a member of the Melitaeini, of which more than half the European members also feed on *P. lanceolata*[26,27]. In southeast France, three of five melitaeines studied were entirely dependent on *P. lanceolata*, and effectively restricted to traditionally managed hay meadows[28]. In the past, human activities provided major ecological and evolutionary opportunities, which the insects were able to exploit because farming was small-scale (butterflies did not need to move far to find fresh habitat) and agricultural developments were slow. Over much of Europe, the Melitaeini (and many other butterflies) became dependent on human activities[29,30]. They are now declining rapidly as a result of the fertilization of ancient hay meadows, and the abandonment of those which would be uneconomic to 'improve'[28,31]. Recent agricultural changes have been so rapid that these insects have been unable to adapt to them. Instead, population after population has become extinct. On present evidence, the Schneider population of *E. editha* is hurrying to join its European relatives in dependence on human maintenance of its host plant and habitat. □

Received 27 May; accepted 22 September 1993.

1. Strong, D. R. *Science* **185,** 1064–1066 (1974).
2. Strong, D. R., Lawton, J. H. & Southwood, T. R. E. *Insects on Plants: Community Patterns and Mechanisms* (Blackwell, Oxford, 1984).
3. Cornell, H. V. & Hawkins, B. A. *Am. Nat.* (in the press).
4. Thomas, C. D., Singer, M. C., Mallet, J. L. B., Parmesan, C. & Billington, H. L. *Evolution* **41,** 892–901 (1987).
5. Singer, M. C. *Symp. R. ent. Soc., Lond.* **11,** 81–88 (1984).
6. Wiklund, C. *Oecologia* **18,** 185–197 (1975).
7. Jermy, T. *Am. Nat.* **124,** 609–630 (1984).
8. Rausher, M. D. *Evolution* **38,** 582–592 (1984).
9. Thompson, J. N. *Evolution* **42,** 118–128 (1988).
10. Futuyma, D. J. & McCafferty, S. S. *Evolution* **44,** 1885–1913 (1990).
11. Jaenike, J. *A. Rev. ecol. Syst.* **21,** 243–273 (1990).
12. Prokopy, R. J., Diehl, S. R. & Cooley, S. S. *Oecologia* **76,** 138–147 (1988).
13. Via, S. *A. Rev. Ent.* **35,** 421–426 (1991).
14. Scriber, J. M., Biebink, B. L. & Snider, D. *Oecologia* **87,** 360–368 (1991).
15. Bowers, M. D., Stamp, N. E. & Collinge, S. K. *Ecology* **73,** 526–536 (1992).
16. Singer, M. C., Ng, D., Vasco, D. & Thomas, C. D. *Am. Nat.* **139,** 9–20 (1992).
17. Grant, B. R. & Grant, P. R. *Proc. R. Soc.* B**251,** 111–117 (1993).
18. Ehrlich, P. R. *Science* **134,** 108–109 (1961).
19. Singer, M. C., Ng, D. & Thomas, C. D. *Evolution* **42,** 977–985 (1988).
20. Moore, S. D. *Ecology* **70,** 1726–1737 (1989).
21. Jaenike, J. *Evol. Ecol.* **7,** 103–108 (1993).
22. Thomas, C. D. & Singer, M. C. *Ecology* **68,** 1262–1267 (1987).
23. Singer, M. C., Thomas, C. D., Billington, H. L. & Parmesan, C. *Anim. Behav.* **37,** 751–759 (1989).
24. Singer, M. C. *Evolution* **37,** 389–403 (1983).
25. Harrison, S., Murphy, D. D. & Ehrlich, P. R. *Am. Nat.* **132,** 360–382 (1988).
26. Higgins, L. G. & Riley, N. D. *A Field Guide to the Butterflies of Britain and Europe* 4th edn (Collins, London, 1980).
27. Chinery, M. *New Generation Guide to the Butterflies and Day-flying Moths of Britain and Europe* (Collins, London, 1989).
28. Warren, M. S. *Bull. Br. ecol. Soc.* **16,** 24–26 (1985).
29. Thomas, J. A. *Symp. R. ent. Soc., Lond.* **11,** 333–353 (1984).
30. Thomas, J. A. *Symp. Br. ecol. Soc.* **31,** 149–197 (1991).
31. Warren, M. S. *Bull. Br. ecol. Soc.* **20,** 212–216 (1989).

ACKNOWLEDGEMENTS. We thank D. Boughton, P. Chippindale, P. R. Ehrlich, A. Fishbach, E. Fleishman, L. Gilbert, A. Launer, R. A. Moore, S. D. Moore, D. D. Murphy, D. Ng, G. Perry, H. Petursson, R. Roberts, E. Vargo and P. J. DeVries. Work at Rabbit Meadow courtesy of US Forest Service (Sequoia National Forest, Hume Lake District) and at Schneider's Meadow courtesy of B., J. & T. Schneider. Work supported by the National Science Foundation, the University of Texas and the authors.

Hippocampal pyramidal cells excite inhibitory neurons through a single release site

A. I. Gulyás*, R. Miles†, A. Sík*, K. Tóth*†, N. Tamamaki‡ & T. F. Freund*§

* Institute of Experimental Medicine, Hungarian Academy of Sciences, Budapest, PO Box 67, H-1450, Hungary
† Laboratoire de Neurobiologie Cellulaire, INSERM U261, Institut Pasteur, 25 rue de Dr Roux, 75264 Paris Cedex 15, France
‡ Department of Anatomy, Fukui Medical School, Matsuoka, Fukui 910-11, Japan

MORPHOLOGICALLY a synapse consists of a presynaptic release site containing vesicles, a postsynaptic element with membrane specialization, and a synaptic cleft between them[1]. The number of release sites shapes the properties of synaptic transmission between neurons[2–4]. Although excitatory interactions between cortical neurons have been examined[5–9], the number of release sites remains unknown. We have now recorded excitatory postsynaptic potentials evoked by single pyramidal cells in hippocampal interneurons and visualized both cells using biocytin injections. Light and electron microscopy showed that excitatory postsynaptic potentials were mediated by a single synapse. We also reconstructed the entire axon arborization of single pyramidal cells, filled *in vivo*, in sections counterstained for parvalbumin, which selectively marks basket and axo-axonic cells[10,11]. Single synaptic contacts between pyramidal cells and parvalbumin-containing neurons were dominant (>80%), providing evidence for high convergence and divergence in hippocampal networks[12].

Transmission often fails (Fig. 1*a*) at excitatory connections between hippocampal neurons[5–7,9]. Failures of synaptic transmission may reflect a low probability of transmitter release, or few release sites[2]. We performed two types of experiments using light and electron microscopy to define the number of transmitter release sites at an excitatory connection: (1) excitatory interactions between hippocampal pyramidal and inhibitory cells were characterized physiologically and their morphological substrate revealed after injecting biocytin in both cells; and (2) we examined contacts between a single filled pyramidal cell axon and inhibitory cells stained immunocytochemically to reveal their calcium-binding protein, parvalbumin[10,11,13].

Simultaneous intracellular records were made from CA3 pyramidal cells and inhibitory cells located close to stratum pyramidale. Inhibitory cells were identified by a distinctive firing pattern[14] by showing that they inhibited nearby pyramidal cells[9], and from their morphology after biocytin injection. Excitatory interactions were detected in ~10% of the recorded pairs (Fig. 1*a*). Excitatory postsynaptic potentials (e.p.s.ps) had a mean amplitude of 0.2–1.5 mV and a rise time from 10–90% of 1.2–2.8 ms ($n=18$). Averaging selected responses to presynaptic action potentials suggested that in 16 of 18 connections some presynaptic action potentials elicited no postsynaptic responses.

Filled cells with complete and reliable light and electron microscopy were recovered from three slices. Inhibitory cells were of different types: one had an axon arborizing in strata lacunosum-moleculare and oriens (Fig. 1*b*), whereas the other two were basket cells innervating predominantly stratum pyramidale. Pyramidal cell axons were intensely stained, and always terminated in a cut end on the slice surface in CA3 or CA1. In two slices, single pyramidal cells were filled; in the third case, two pyramidal cells were recovered, although electrical stimulation appeared to generate single action potentials.

Each axon established a single contact on a second-order inhibitory cell dendrite at 80–250 μm from the soma (arrow in Fig. 1*b*). Electron microscopy confirmed that each contact was a conventional synapse (Fig. 2*b*). Presynaptic elements contained round clear vesicles and serial sectioning revealed a single synaptic active zone of diameter 0.4–0.8 μm.

§ To whom correspondence should be addressed.

Quantal amplitude (q) and probability of transmitter release (p) were evaluated from e.p.s.p. amplitude distributions. For the interaction shown in Fig. 1, q was estimated as 0.65 ± 0.25 mV and p was 0.37, with $N=1$ corresponding to the single anatomically defined site. Figure 2*d* shows the resulting fit of the e.p.s.p. amplitude distribution. Statistical analysis suggested 14 of the other 17 interactions involved a single release site, including two cases with confirmatory anatomy (Fig. 3). In three cases, slightly better fits were obtained with $N=2$ (1) and $N=3$ (2).

Mean quantal amplitude was estimated as 0.79 ± 0.20 mV ($n=18$). We estimate that the coefficient of variation of quantal amplitude was 21–53% (mean, $35 \pm 9\%$; $n=18$). This compares with values of 40–50% measured for miniature e.p.s.cs in cortical and hippocampal pyramidal cells when variance from electrotonic sources was minimized[15,16]. The mean value derived for probability of transmitter release was 0.75 ± 0.19. In four cases, p was higher than 0.9 (Fig. 2*e*). Similar high values for p have been inferred for some excitatory terminals on spinal motor neurons[3]. However, our estimate may be high because experiments were done at high calcium concentrations, which enhances transmitter release, and because larger, more frequently evoked e.p.s.ps were more readily seen.

The hypothesis that pyramidal cells make single synaptic contacts with inhibitory cells was tested further by filling single CA3 pyramidal cells with neurobiotin *in vivo*[17]. Parvalbumin-

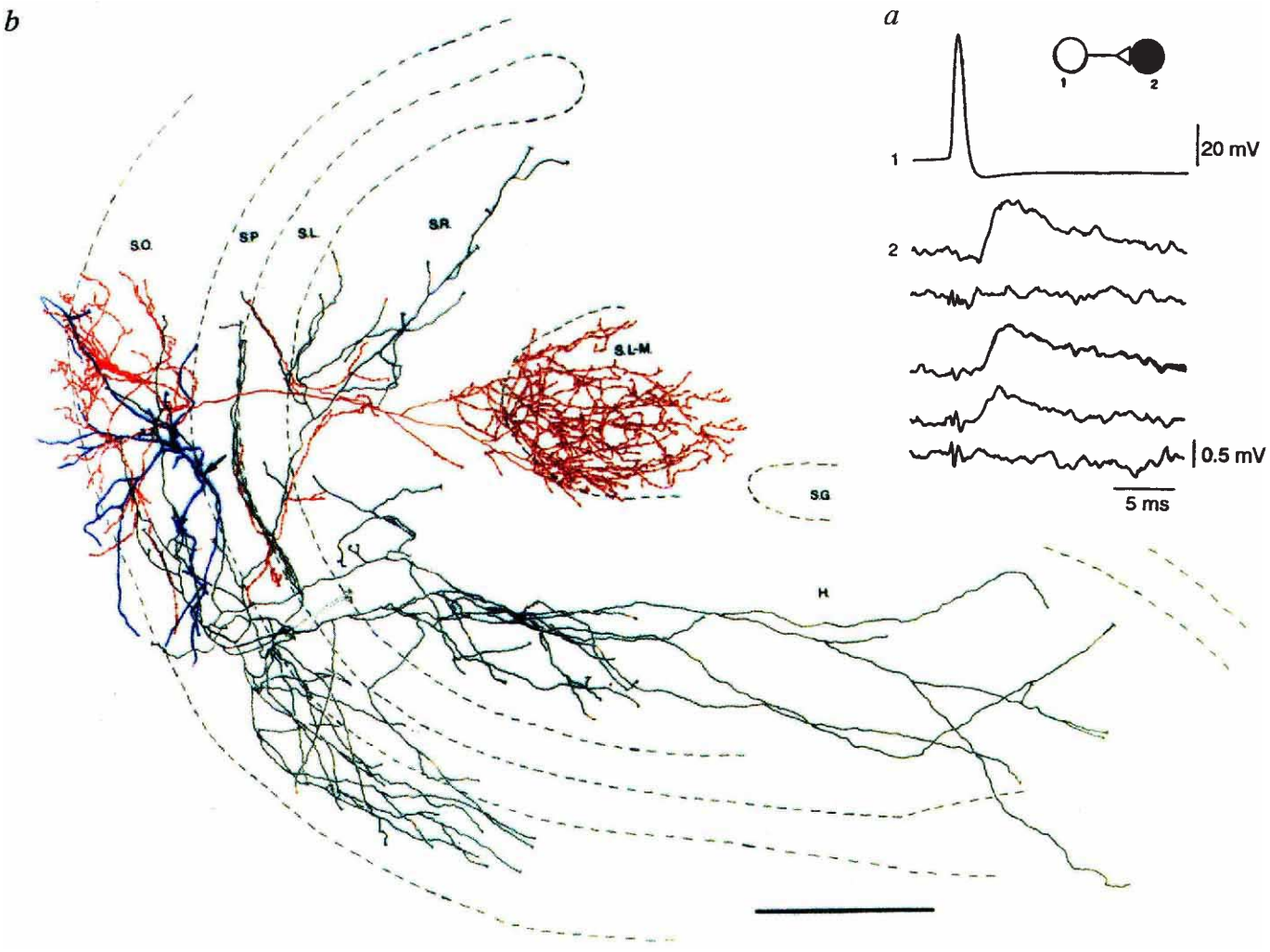


FIG. 1 Synaptic transmission between a pyramidal and an inhibitory cell in the CA3 region of the hippocampus. *a*, Single pyramidal cell action potentials (1) evoke e.p.s.ps in the inhibitory cell (2). E.p.s.ps fluctuated in amplitude and transmission sometimes failed (second and fifth traces). *b*, Camera lucida reconstruction of the axonal arbor of the presynaptic pyramidal cell (black), and the dendritic tree (blue) and axonal arbor (red) of the postsynaptic inhibitory cell. The pyramidal cell axon established a single contact on a mid-distal dendrite of the inhibitory cell in stratum oriens (arrow). Scale, 200 μm. Abbreviations: SO, stratum oriens; SP, str. pyramidale; SL, str. lucidum; SR, str. radiatum; SL-M, str. lacunosum-moleculare; SG, str. granulosum; H, hilus of the dentate gyrus.

METHODS. Slices of hippocampus (thickness 450 μm) from guinea-pigs were prepared and maintained as before[9], except that animals were perfused through the heart under ether anaesthesia with ice-cold physiological saline to remove red blood cells. The saline contained (in mM) NaCl, 128; KCl, 3; $NaHCO_3$, 26; $CaCl_2$, 4; $MgCl_2$, 4; D-glucose, 10. Recording electrodes contained 4% biocytin in 0.5 M potassium-acetate and had a resistance of 40–80 MΩ. After an excitatory synaptic interaction was characterized, biocytin was injected iontophoretically (2 nA, 300 ms, 1 Hz, for 10–20 min) into both cells. Slices were then kept in the recording chamber for 60–90 min before overnight fixation in 4% paraformaldehyde, 0.05% glutaraldehyde and 15% picric acid in 0.1 M phosphate buffer. They were sectioned on a vibratome at 80 μm and freeze–thawed 3 times above liquid nitrogen in 0.1 M phosphate buffer containing 10% glycerol and 15% sucrose. Injected neurons were visualized using the avidin–biotinylated horse-radish peroxidase complex reaction (Vector Labs; Elite ABC kit) with nickel-intensified 3-3′-diaminobenzidine (Sigma) as chromogen. Sections were processed for electron microscopy as described[21]. Axons and dendrites of filled cells were reconstructed with a camera lucida from the light microscope. Potential synaptic contacts were then re-embedded and sectioned for electron microscopy.

containing inhibitory cells (basket and axo-axonic cells) were visualized immunocytochemically in the same sections from the entire hippocampus (Fig. 4). This pyramidal cell axon was followed through 45 sections (60 μm thick) ramifying in the CA3 and CA1 regions and extending 2.7 mm along the longitudinal axis. Figure 4*b* plots locations from nine representative consecutive sections, where axon varicosities contact parvalbumin-positive inhibitory cell dendrites (134) or soma (1). For 93% of potential target neurons (125 out of 134), a single contact was formed by the filled axon. In another study, 85% of potential pyramidal cell contacts with parvalbumin-containing neurons involved a single bouton[18]. Electron microscopy showed that 81% of these potential contacts from light microscopy were indeed synapses with one active zone. Convergence of multiple pyramidal cell axon collaterals onto different dendrites of one parvalbumin-positive cell was not observed.

These data show that pyramidal cells form single axonal contacts with postsynaptic inhibitory neurons. This arrangement differs from other connections such as climbing fibre synapses onto cerebellar Purkinje cells, where there are 200–300 transmitter release sites[19]. In the hippocampus, morphological studies suggest that inhibitory cells form 5–30 synaptic contacts with a

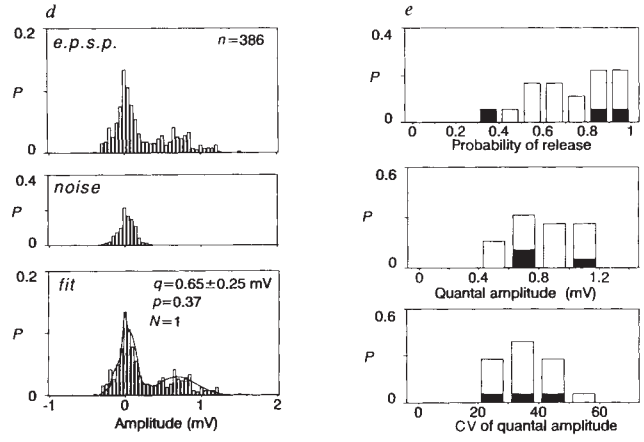


FIG. 2 Light and electron microscopic substrates of synaptic transmission. Light (*a*) and electron micrographs (*b*) of the contact shown in Fig. 1. Pre- and postsynaptic specializations and a widening of the synaptic cleft between the pyramidal cell axon terminal (b_1) and the inhibitory cell dendrite are clearly visible (arrows). *c*, An adjacent bouton (b_2, also seen in *a*) of the labelled axon established an asymmetrical synapse (arrow) on a dendritic spine, confirming that the axon arose from the pyramidal cell. Scale in *a*, 10 μm; in *b* and *c*, 0.5 μm. *d*, Distribution of e.p.s.p. amplitudes (measured from baseline to peak) and noise (measured with the same time difference from baseline to baseline) for 386 events elicited at 1 Hz. (*P*, relative probability of events.) During data collection, postsynaptic membrane potential was maintained within ±3 mV of a value between −60 and −70 mV. Fits to amplitude histograms were made by constructing distributions with *N*, the number of transmitter release sites, *p*, the probability of release, and *q*, the postsynaptic effect of a single quantum[9]. Transmission was assumed to occur in quantal steps (*q*, 2*q*, 3*q* and so on) and *p* followed binomial statistics (that is, release from different sites was independent). The parameter *q* could be fixed or could have a coefficient of variation (CV) δq. These theoretical distributions were convolved with experimentally determined noise distributions and compared with those of e.p.s.p. amplitudes. Fits were generated by minimizing the sum of the squared difference between the distributions. Best fits were obtained with $N=1$ for the 3 experiments in which a single release site was anatomically verified. We obtained a value of 0.37 for *p*, and 0.65 ± 0.25 mV for *q* in the interaction in *a* and *b*. For the 15 interactions without anatomical data, $N=1$ gave best fits in 11 cases. Differences in the sum of squares $N=1$ and $N=2$ ranged between 1 and 75% (mean difference 15%, $n=15$). $N=2$ and $N=3$ agreed better with experimental data in 1 and 2 interactions respectively. In these cases, the minimum sum of squares was 1–3% better for $N=2$ or 3 than for $N=1$. Estimated values for *p*, *q* and δq are shown in *e*. Values from experiments with an anatomically verified single release site are shown in black.

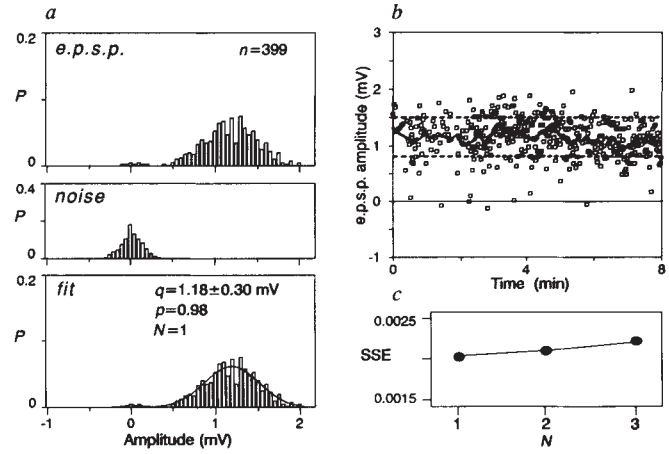


FIG. 3 Fluctuation of e.p.s.p. amplitude at a connection with a single morphologically identified transmitter release site and a high probability of release. *a*, Distribution of e.p.s.p. amplitudes and corresponding noise measurements for 399 events evoked during 8 min of presynaptic stimulation at 1 Hz. Note that not all current injections elicited an action potential. The fit was made with $N=1$, $p=0.98$ and $q=1.18\pm0.30$ mV ($\delta q=25\%$). *b*, Stationarity of data. The amplitudes of e.p.s.ps from which the histogram was constructed are plotted against time. The solid line is a running average of 20 events. The mean amplitude of the first 50 responses was 1.15 mV and the dotted lines are plotted at ±20% of this value. Data sets were constructed from periods when such running averages were maintained within 30% of the initial mean e.p.s.p. amplitude. *c*, Goodness of fit. The sum of squares of differences (SSE) between experimental data (e.p.s.p.) and fits obtained with $N=1$, $N=2$ ($P=0.99$, $q=0.59\pm0.28$ mV) and $N=3$ ($P=0.91$, $Q=0.42\pm0.22$ mV).

single pyramidal cell[20,21], indicating that there may be different strategies for inhibitory and excitatory transmission.

E.p.s.ps initiated in interneurons by single pyramidal cell action potentials are mediated by a single morphological synapse. Thus transmission failures and release probabilities reflect release from a single site, and variation between successive postsynaptic events is variation occurring at one synaptic complex. It is thought that only one vesicle is released from a single presynaptic site[2-4,16,22], although this has not been directly verified. If so, transmission at this synapse would be mono-quantal, as suggested for connections made by CA3 pyramidal cells on CA1 pyramidal cells[23,24]. Mono-quantal transmission would constrain mechanisms of synaptic plasticity. If at a functional connection there is only one morphological transmitter release site, then latent release sites cannot be uncovered. Presynaptic plasticity must then be limited to changes in release probability[25], which should saturate, or to changes in the amount of transmitter released from one site[16].

Our results also provide numerical values for connectivity needed to model neuronal networks[12,26,27]. As over 1,000 excitatory synapses terminate on a single inhibitory cell[28], our data suggest that more than 1,000 pyramidal cells converge onto one

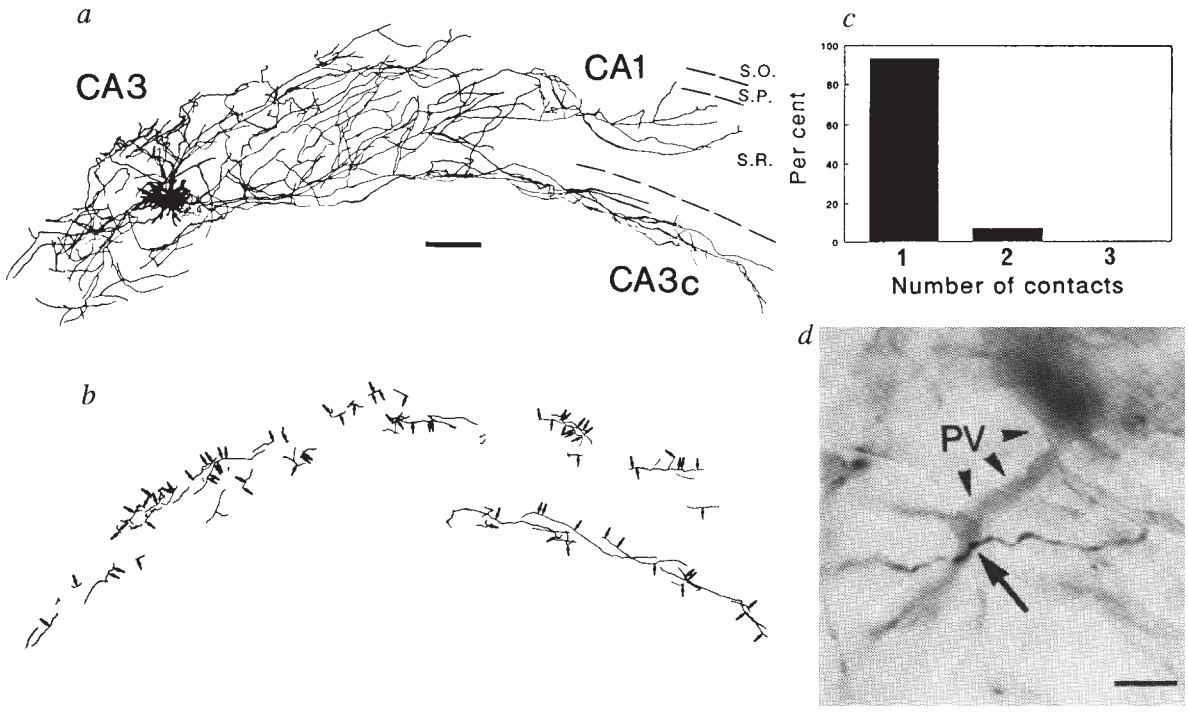


FIG. 4 Synaptic contacts between a single pyramidal cell and the population of postsynaptic parvalbumin-immunoreactive inhibitory cells. *a*, The reconstructed axonal arborization of a neurobiotin-labelled CA3 pyramidal cell from the rat hippocampus in one hemisphere. The axon spread through the stratum radiatum (SR) and oriens (SO) of 43 60-μm-thick sections (2.6 mm). Axons from sections 1–27 are shown. *b*, Even spatial distribution of pyramidal cell terminals contacting parvalbumin (PV)-positive dendrites (arrows) taken from 9 sections. *c*, Histogram of the number of contacts at pyramidal cell/parvalbumin-positive cell junctions. *d*, Light micrograph of a contact between a pyramidal cell axon terminal (arrow) and parvalbumin-positive inhibitory cell dendrite (arrowheads). Scale in *a* is 200 μm, and 10 μm in *d*.

METHODS. Cells were filled *in vivo* with neurobiotin (10% in 0.5 M KCl) by pressure application[17] and visualized as for cells from slices. Neurons containing parvalbumin were immunostained in the same sections using the peroxidase–antiperoxidase technique and a well characterized antiserum[10].

inhibitory neuron. About 2% of CA3 pyramidal cell targets are immunoreactive for parvalbumin. Thus, a pyramidal cell possessing 10–50 thousand boutons[18,29] may excite 200–1,000 parvalbumin-containing inhibitory cells. E.p.s.ps initiated by single pyramidal cells may cause inhibitory cell firing[9]. This excitatory drive presumably contributes to the high firing frequency of many hippocampal interneurons[11,14]. Thus the activation of a single CA3 pyramidal cell may, owing to high divergence in excitatory and inhibitory connections, disynaptically inhibit thousands of pyramidal cells in the CA3 and CA1 regions. This circuitry will operate in some, but not all, behavioural states to exert a strong, widespread inhibitory control of pyramidal cell firing[30]. □

Received 1 June; accepted 5 October 1993.

1. Heuser, J. E. & Reese, T. S. *Handbook of Physiology—The Nervous System* Vol. 1 (ed. Kandel, E. R.) 261–294 (Am. Phys. Soc., Bethesda, 1977).
2. Korn, H. & Faber, D. S. *Synaptic Function* (eds Edelman, G. M., Gall, W. E. & Cowan, W. M.) 57–108 (Wiley, New York, 1987).
3. Redman, S. *Physiol. Rev.* **70,** 165–198 (1990).
4. Stevens, C. F. *Neuron* **10,** 55–64 (1993).
5. Miles, R. & Wong, R. K. S. *J. Physiol.* **373,** 397–418 (1986).
6. Sayer, R. J., Redman, S. & Andersen, P. *J. Neurosci.* **9,** 840–850 (1989).
7. Malinow, R. *Science* **252,** 722–724 (1991).
8. Stern, P., Edwards, F. A. & Sakmann, B. *J. Physiol.* **449,** 247–278 (1992).
9. Miles, R. *J. Physiol.* **428,** 61–77 (1990).
10. Baimbridge, K. G., Miller, J. J. & Parkes, C. O. *Brain Res.* **239,** 519–525 (1982).
11. Kawaguchi, Y., Katsumaru, H., Kosaka, T., Heizmann, C. W. & Hama, K. *Brain Res.* **416,** 369–374 (1987).
12. Traub, R. D. & Miles, R. *Neuronal Networks of the Hippocampus* (Cambridge Univ. Press, UK, 1991).
13. Kosaka, T., Katsumaru, H., Hama, K., Wu, J. J. & Heizmann, C. W. *Brain Res.* **419,** 119–130 (1987).
14. Schwartzkroin, P. A. & Mathers, L. H. *Brain Res.* **157,** 1–10 (1978).
15. Hestrin, S. *Neuron* **9,** 991–999 (1993).
16. Bekkers, J. M., Richerson, G. B. & Stevens, C. F. *Proc. natn. Acad. Sci. U.S.A.* **87,** 5359–5362 (1990).
17. Tamamaki, N., Watanabe, K. & Nojyo, Y. *Brain Res.* **307,** 336–340 (1984).
18. Sik, A., Tamamaki, N. & Freund, T. F. *Eur. J. Neurosci.* (in the press).
19. Llinas, R., Bloedel, J. R. & Hillman, D. E. *J. Neurophysiol.* **32,** 847–870 (1969).
20. Somogyi, P., Kisvarday, Z. F., Martin, K. A. C. & Whitteridge, D. *Neuroscience* **10,** 261–294 (1983).
21. Freund, T. F., Martin, K. A. C., Somogyi, P. & Whitteridge, D. *J. comp. Neurol.* **242,** 275–291 (1985).
22. Busch, C. & Sakmann, B. *Cold Spring Harbor Symp.* **55,** 69–80 (1990).
23. Manabe, T., Renner, P. & Nicoll, R. A. *Nature* **355,** 50–55 (1992).
24. Raastad, M., Storm, J. F. & Andersen, P. *Eur. J. Neurosci.* **4,** 113–117 (1992).
25. Malgaroli, A. & Tsien, R. W. *Nature* **357,** 134–139 (1992).
26. McClelland, J. L. *et al. Parallel Distributed Processing* (MIT Press/Bradford Books, Massachusetts, 1986).
27. McNaughton, B. L. & Morris, N. G. *Trends Neurochem. Sci.* **10,** 408–415 (1987).
28. White, E. L. & Keller, A. *J. comp. Neurol.* **262,** 13–26 (1987).
29. Li, X. G., Somogyi, P., Ylinen, A. & Buzsáki, G. *J. comp. Neurol.* (in the press).
30. Buzsáki, G., Leung, L. W. S. & Vanderwolf, C. H. *Brain Res. Rev.* **6,** 139–171 (1983).

ACKNOWLEDGEMENTS. We thank G. Buzsáki for discussion and for comments on the manuscript, K. G. Baimbridge for antisera against parvalbumin, and E. Borók, I. Weisz, C. Pauletti and G. Terstyánszky for technical assistance. This work was supported by the Human Frontier Science Program Organization, the European Science Foundation, the Ministère de la Recherche et de la Technologie, the Hungarian Science Foundation, and OTKA Hungary.

Role of a metabotropic glutamate receptor in synaptic modulation in the accessory olfactory bulb

Yasunori Hayashi*†, Akiko Momiyama‡, Tomoyuki Takahashi‡§, Hitoshi Ohishi‖, Reiko Ogawa-Meguro‖, Ryuichi Shigemoto‖, Noboru Mizuno‖ & Shigetada Nakanishi*¶

* Institute for Immunology, † Department of Pharmacology, ‡ Department of Physiology, ‖ Department of Morphological Brain Science, Kyoto University Faculty of Medicine, Kyoto 606, Japan

VARIOUS functions of glutamate transmission are mediated by both ionotropic and metabotropic glutamate receptors[1]. The metabotropic glutamate receptors (mGluRs) consist of at least six different subtypes that are classified into three subgroups, mGluR1/mGluR5, mGluR2/mGluR3, and mGluR4/mGluR6 (refs 1–5), but their physiological roles are largely unknown. Here we report the identification of a very potent agonist for mGluR2/mGluR3, DCG-IV, and the specific localization of mGluR2 in granule cell dendrites that form dendrodendritic synapses with mitral cells in the accessory olfactory bulb. Using the DCG-IV agonist for mGluR2 in combination with slice patch-recording, we demonstrate that the granule cell mGluR2 presynaptically suppresses inhibitory GABA (γ-aminobutyrate) transmission to the mitral cell. Our results indicate that mGluR2 in granule cells plays an important role in the persistent excitation of olfactory sensory transmission in the accessory olfactory bulb by relieving mitral cells from the GABA inhibition.

Our previous study indicated that two extended L-isomers of 2-(carboxycyclopropyl)glycines (L-CCGs) are potent and selective agonists for the mGluR family[6]. An L-CCG derivative, (2S,1′R,2′R,3′R)-2-(2,3-dicarboxycyclopropyl)glycine (DCG-IV; Fig. 1) has recently been reported[7,8]. We determined agonist potencies of DCG-IV by examining its effects on the signal transduction of mGluR1, mGluR2, mGluR3 and mGluR4 expressed in Chinese hamster ovary (CHO) cells (Fig. 1). DCG-IV was a very potent agonist for both mGluR2 and mGluR3, with half-maximal effective concentrations (EC_{50}) of 3×10^{-7} M and 2×10^{-7} M, respectively. Remarkably, this compound had no obvious agonist activity for either mGluR1 or mGluR4. DCG-IV could bind to the NMDA (*N*-methyl-D-aspartate) receptor ($EC_{50}=3\times10^{-6}$ M) (Fig. 1) and induced an NMDA response ($EC_{50}\approx3\times10^{-5}$ M) in *Xenopus* oocytes expressing the cloned NMDA receptor (data not shown). Thus, the potency of DCG-IV for the NMDA receptors was at least one order of magnitude lower than that for the activation of mGluR2. DCG-IV had no antagonist activity on any mGluR subtypes, and did not interact with the α-amino-3-hydroxy-5-methyl-4-isoxazolepropionate-(AMPA)-kainate receptors. DCG-IV is thus a very useful agonist for mGluR2/mGluR3.

In situ hybridization analysis showed high expression of mGluR2 messenger RNA in limited regions of the rat brain[9] and particularly in granule cells of the accessory olfactory bulb (AOB) (Fig. 2*a*, *b*). We investigated the subcellular localization of mGluR2 in granule cells by using polyclonal antibody raised against the mGluR2 protein expressed in *Escherichia coli*. On western blot analysis, the mGluR2 antibody reacted with both mGluR2 and mGluR3 expressed individually in COS cells (Fig. 2*d*). This antibody gave rise to two bands representing mGluR2 and mGluR3 in membrane preparations from the cerebral cortex where mGluR3 mRNA is highly expressed[10]. In contrast, only the band corresponding to mGluR2 was observed in the AOB membrane preparation. mGluR3 mRNA is moderately expressed in mitral cells but not appreciably in granule cells of the AOB (ref. 10), indicating that the antibody crossreactivity with mGluR3 is weak. The mGluR2 antibody is thus useful for the subcellular localization of mGluR2 in granule cells.

In the AOB, mitral cells receive afferent inputs from the vomeronasal nerve and transmit excitatory outputs to various brain regions (Fig. 3*a*). Granule cells are inhibitory interneurons that form typical dendrodendritic synapses with mitral cells[11–13]. These synapses mediate reciprocal transmission in which the granule cell is excited by glutamate from the mitral cell and exerts an inhibition onto the mitral cell by GABA (refs 11–13).

§ Present address: Department of Neurophysiology, Institute for Brain Research, Faculty of Medicine, University of Tokyo, Tokyo 113, Japan.
¶ To whom correspondence should be addressed.

In electron microscopic analysis, the glutamatergic synapse can be seen as an asymmetric synapse where dense filamentous materials are located beneath the granule cell synaptic membrane and are confronted with spherical synaptic vesicles of the mitral cell[12,13]. The GABAergic synapse is seen as a symmetrical synapse where pleomorphic vesicles are localized under the membrane of the granule cell dendrite[12,13]. In immunoelectron microscopic analysis with the mGluR2 antibody, the localization of immunoreactivity of mGluR2 in granule cell dendrites was clearly seen (Fig. 2*c*). Furthermore, an immunopositive dendritic profile of a granule cell formed a synaptic contact with an immunonegative dendritic profile of a mitral cell (Fig. 2*c*).

On the basis of the above findings, we investigated the role of

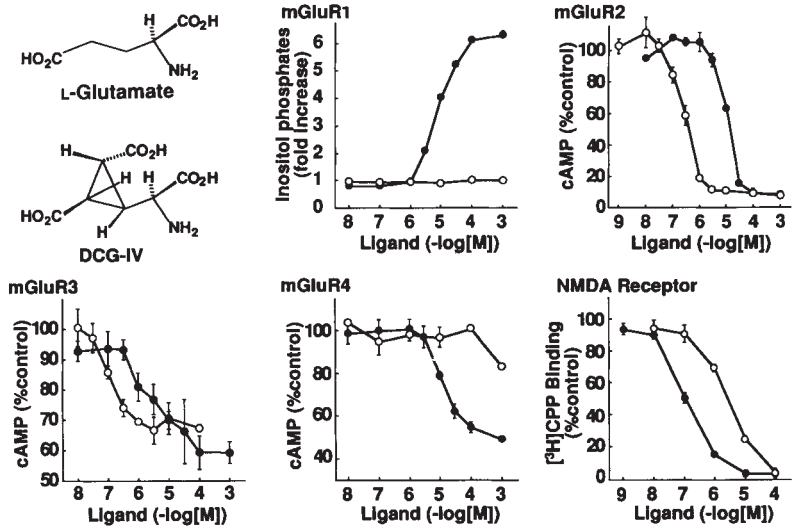


FIG. 1 Potencies and selectivity of DCG-IV for glutamate receptors. The total inositol phosphate formation was determined by incubating mGluR1-expressing CHO cells with test reagents for 20 min (ref. 6) and is expressed as multiples of inositol phosphate levels in agonist-untreated cells. Dose–response curves for inhibition of the forskolin-stimulated cAMP formation were determined by incubating mGluR2-, mGluR3- and mGluR4-expressing cells with test reagents in the presence of 10 μM forskolin for 10 min (ref. 6). cAMP levels in agonist-untreated cells are taken as 100%. Displacement curves of binding of [^{3}H]3-((±)-2-carboxypiperazin-4-yl)propyl-1-phosphate ([^{3}H]CPP, 10 nM) to the NMDA receptor were determined with crude synaptic membranes from adult rat forebrain in the absence (total binding) or presence (nonspecific binding) of 1 mM L-glutamate for 15 min on ice[17]. No appreciable binding of DCG-IV to the AMPA and kainate receptors were observed by similar displacement experiments using [^{3}H]AMPA (5 nM) and [^{3}H]kainate (2 nM), respectively[17]. ●, L-glutamate; ○, DCG-IV. Error bars are shown as the mean ± s.e.

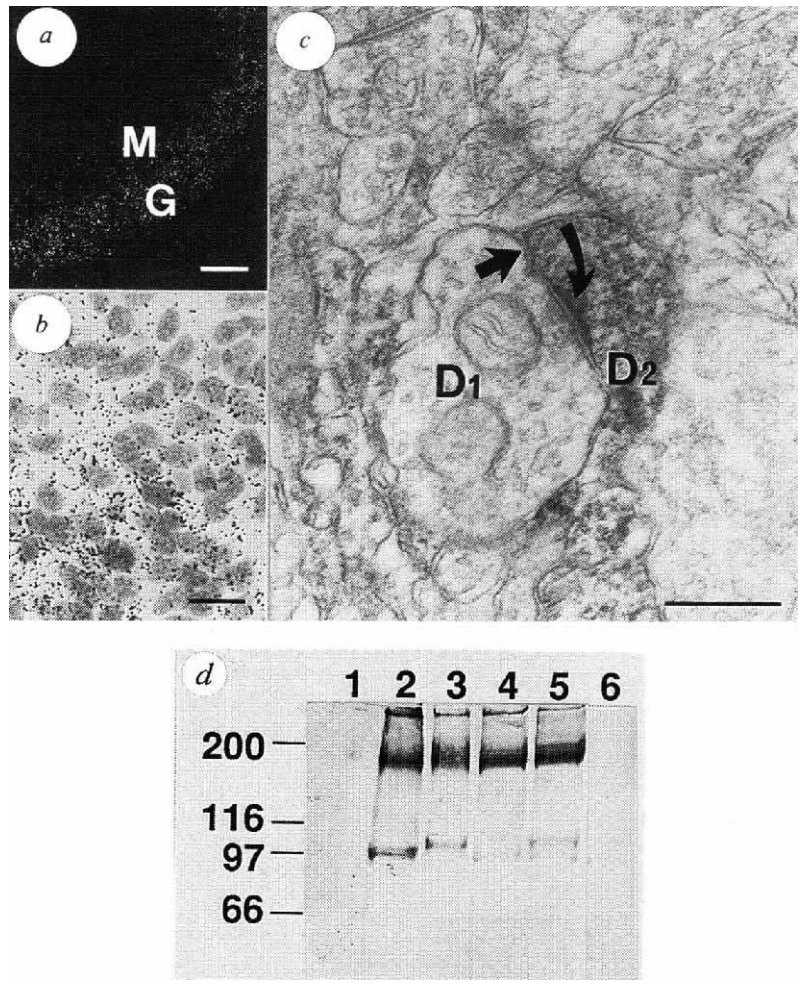


FIG. 2 Subcellular localization of mGluR2 in granule cells of the AOB. Dark- and bright-field photomicrographs of an emulsion-dipped section through the AOB of an 8-day-old rat are indicated in *a* and *b*, respectively. M, Mitral cell layer; G, granule cell layer. An electron micrograph (*c*) shows a dendrodendritic synapse between an immunonegative profile of a mitral cell dendrite (D_1) and an immunopositive profile of a granule cell dendrite (D_2) in the AOB of an adult rat. The D_2 is postsynaptic in an asymmetrical synapse (the curved arrow), and presynaptic in a symmetrical synapse (the straight arrow). Scale bars, 100 μm (*a*), 20 μm (*b*) and 0.5 μm (*c*). *d*, Immunoblot analysis of mGluR2 (~98K) and mGluR3 (~106K) was done with membrane fractions from the following sources: lane 1, non-transfected COS cells; lane 2, mGluR2-expressing COS cells; lane 3, mGluR3-expressing COS cells; lanes 4 and 6, AOB; and lane 5, cerebral cortex. In lane 6, the antibody was preabsorbed with a synthetic C-terminal peptide representing the antigen epitope of mGluR2. Positions of molecular mass (K) markers are indicated on the left.

METHODS. *In situ* hybridization was done using an *in vitro* synthesized, ^{35}S-labelled RNA probe corresponding to the 616-base-pair *Apa*I–*Sac*I fragment of the mGluR2 cDNA clone as described[2]. For the immunoelectron microscopic and immunoblot analyses, the construction, overexpression, purification and antibody production of the *trpE*-mGluR2 fusion protein were done as described[18]. The *trpE*-mGluR2 fusion protein containing a part of transmembrane segment VII and its following C-terminal portion of mGluR2 was made by inserting the 229-base-pair *Fok*I fragment of the mGluR2 cDNA into the *Eco*RI site of a pATH3 expression vector[18]. Immunoelectron microscopy of an AOB and immunoblot analysis of mGluR2 and mGluR3 were done using the affinity-purified antibody raised against mGluR2 according to the procedures described[18]. The higher-molecular-mass components probably represent polymeric aggregates of the receptor, because they were not observed with the preabsorbed antibody and were also stained inconsistently from experiment to experiment.

mGluR2 in synaptic transmission between the mitral cell and granule cell by examining the effect of DCG-IV on the GABA transmission to the mitral cell. Whole-cell recording was done from a mitral cell visually identified in thin slices at a holding potential of −70 mV (refs 14, 15) (Fig. 3*a*). The extracellular stimulation of a granule cell evoked an inward current (Fig. 3*b*). This response was blocked by the $GABA_A$-receptor antagonist bicuculline in a reversible manner (Fig. 3*b*). Also, when a mitral cell was electrically stimulated, a granule cell elicited excitatory postsynaptic currents that were sensitive to the AMPA/kainate receptor antagonist 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX) (data not shown). We thus confirmed that the mitral

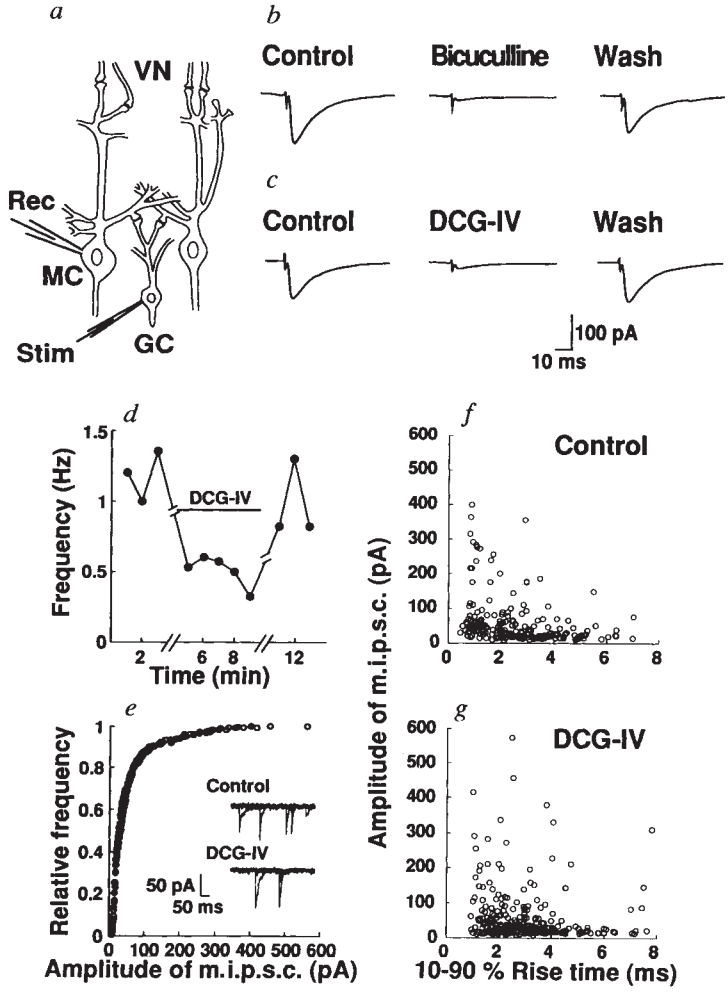


FIG. 3 Suppression of GABA-mediated i.p.s.cs by DCG-IV. As shown in *a*, whole-cell recording (Rec) of GABA-mediated i.p.s.cs was conducted by patch clamping of a cell body of a mitral cell (MC), and extracellular stimuli (Stim) were applied to a granule cell (GC) forming dendrodendritic synapses with the recorded mitral cell; VN, vomeronasal nerve. In *b* and *c*, current traces of GABA-mediated i.p.s.cs were recorded in the presence of 10 μM bicuculline and 3 μM DCG-IV, respectively. All traces are averaged records of 60 trials. Left traces, before drug application; middle traces, 90 s after drug application; right traces, after washing. The effect of DCG-IV on the frequency of m.i.p.s.cs is indicated in *d*. Each point represents the mean frequency of m.i.p.s.cs in a 1-min period. *e*, Cumulative probability distributions of amplitudes of spontaneous m.i.p.s.cs are plotted by recording mitral cells before (●; 219 events) and after application of 3 μM DCG-IV (○; 288 events), respectively. Miniature i.p.s.cs are shown in insets with 5 sweeps superimposed. There was no statistically significant difference in the two plots as assessed by the Kolmogorov–Smirnov test[19]. The rise time (10–90%)–amplitude relationships of m.i.p.s.cs before and after DCG-IV application are indicated in *f* and *g*, respectively. The mean rise times were 2.57 ms in the control and 2.93 ms in DCG-IV treatment.

METHODS. Coronal slices 130 μm-thick were prepared from 6–8-day-old rats using a tissue slicer in an oxygenated and chilled Krebs solution. Whole-cell recording from a mitral cell and stimulation of a granule cell were done by the slice patch-recording method[14,15]. The recording chamber was continuously perfused with a solution of the following compositions: 113 mM NaCl; 3 mM KCl; 1 mM NaH_2PO_4; 25 mM $NaHCO_3$; 11 mM glucose; 2 mM $CaCl_2$; 1 mM $MgCl_2$ (pH adjusted to 7.4 by bubbling with 95% O_2/5% CO_2); during recording, 0.5 μM strychnine and 10 μM CNQX were included in the perfusate[20]. The recording pipette contained a solution of the following compositions: 140 mM CsCl; 9 mM NaCl; 1 mM $MgCl_2$; 1 mM EGTA; 10 mM HEPES (pH adjusted to 7.3 with KOH). The pipette for stimulation was filled with 1 M NaCl and placed on a single granule cell. Miniature i.p.s.cs were recorded from mitral cells as described except that 0.3 μM tetrodotoxin was added to the perfusate.

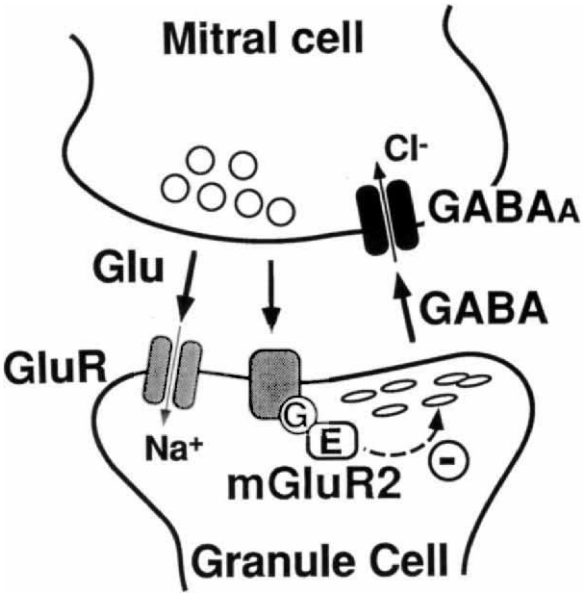


FIG. 4 A model of a modulatory role of mGluR2 in dendrodendritic synaptic transmission between a granule cell and a mitral cell. For detailed explanation, see text. The inhibition of GABA release is marked by a minus symbol but its mechanism remains unknown. GluR, ionotropic glutamate receptor; $GABA_A$, $GABA_A$ receptor; G, G protein; E, intracellular effector.

cells and the granule cells mediate reciprocal synaptic transmission. When DCG-IV was added to the perfusate, the amplitude of the GABA-mediated inhibitory postsynaptic current (i.p.s.c.) was markedly reduced ($28.4 \pm 15.3\%$ of control levels; $n = 5$) in a reversible manner (Fig. 3*c*). At this drug concentration, no appreciable change was observed for the holding current. In this experiment, i.p.s.cs were measured in the perfusate containing 1 mM Mg^{2+} which blocks the NMDA receptors[16]. In addition, i.p.s.cs were inhibited to a similar extent ($35.4 \pm 26.3\%$ of control levels; $n = 5$) in the presence of an NMDA receptor antagonist (50 μM D,L-2-amino-5-phosphonovalerate) (data not shown). Thus, the suppression of i.p.s.cs is not due to the action of DCG-IV on the NMDA receptors. We also confirmed that DCG-IV had no effect on mitral cell responses elicited by application of the $GABA_A$ receptor agonist muscimol (0.5 μM), and L-2-amino-4-phosphonobutyrate (50 μM), a potent mGluR4 subgroup agonist, had no effect on electrically evoked i.p.s.cs in mitral cells (data not shown). These results demonstrate that the activation of mGluR2 presynaptically suppresses a GABA inhibition on the mitral cell.

The presynaptic action of DCG-IV was further confirmed electrophysiologically by analysing spontaneous miniature i.p.s.cs (m.i.p.s.cs). DCG-IV suppressed the mean frequency of m.i.p.s.cs ($62.7 \pm 19.7\%$ of control) without any change in either the cumulative probability of the amplitude distribution of m.i.p.s.cs or the rise time–amplitude relationship of m.i.p.s.cs before and after application of DCG-IV ($n = 3$) (Fig. 3*d–g*). In contrast, m.i.p.s.cs almost completely disappeared after application of bicuculline (data not shown). These results indicate that DCG-IV presynaptically reduces GABA release without changing postsynaptic responsiveness to GABA.

This study demonstrates that mGluR2 is present at the granule cell dendrites and presynaptically suppresses inhibitory GABA transmission to the mitral cell. Glutamate released from an excited mitral cell not only stimulates granule cells through the activation of the ionotropic glutamate receptor but also relieves the GABA-mediated inhibition onto the excited mitral cell through stimulation of mGluR2, as illustrated schematically in Fig. 4. The granule cell lacks an axon and forms divergent dendrodendritic synapses with a large number of surrounding mitral cells[11–13] (see also Fig. 3*a*). The excitation of mitral cells thus evokes lateral inhibition in the neighbouring mitral cells through trans-synaptically excited granule cells[11–13]. Under the mechanism revealed in this study, however, it can be postulated that the GABA inhibition is relieved in an excited mitral cell by the activation of mGluR2. Furthermore, this activation is thought to be confined to the synapses of the excited mitral cells and would thus maintain the lateral inhibition of unexcited neighbouring mitral cells. This mechanism would evidently enhance the signal-to-noise ratio between the excited mitral cells and their neighbouring mitral cells. The mGluR2-mediated modulation in the microcircuitry between the mitral cell and granule cell may be important in discrimination and resolution of the olfactory sensory transmission in the AOB. □

Received 29 June; accepted 6 October 1993.

1. Nakanishi, S. *Science* **258,** 597–603 (1992).
2. Masu, M., Tanabe, Y., Tsuchida, K., Shigemoto, R. & Nakanishi, S. *Nature* **349,** 760–765 (1991).
3. Houamed, K. M. *et al. Science* **252,** 1318–1321 (1991).
4. Tanabe, Y., Masu, M., Ishii, T., Shigemoto, R. & Nakanishi, S. *Neuron* **8,** 169–179 (1992).
5. Abe, T. *et al. J. biol. Chem.* **267,** 13361–13368 (1992).
6. Hayashi, Y. *et al. Br. J. Phamac.* **107,** 539–543 (1992).
7. Ohfune, Y., Shimamoto, K., Ishida, M. & Shinozaki, H. *Bioorg. med. Chem. Lett.* **3,** 15–18 (1993).
8. Ishida, M., Saitoh, T., Shimamoto, K., Ohfune, Y. & Shinozaki, H. *Br. J. Pharmac.* **109,** 1169–1177 (1993).
9. Ohishi, H., Shigemoto, R., Nakanishi, S. & Mizuno, N. *Neuroscience* **53,** 1009–1018 (1993).
10. Ohishi, H., Shigemoto, R., Nakanishi, S. & Mizuno, N. *J. comp. Neurol.* **335,** 252–266 (1993).
11. Rall, W. & Shepherd, G. M. *J. Neurophysiol.* **31,** 884–915 (1968).
12. Mori, K. *Prog. Neurobiol.* **29,** 275–320 (1987).
13. Shepherd, G. M. *Sci. Am.* **238,** 92–103 (1978).
14. Edwards, F. A., Konnerth, A., Sakmann, B. & Takahashi, T. *Pflügers Arch.* **414,** 600–612 (1989).
15. Takahashi, T. *J. Physiol., Lond.* **450,** 593–611 (1992).
16. Moriyoshi, K. *et al. Nature* **354,** 31–37 (1991).
17. Kawai, M., Horikawa, Y., Ishihara, T., Shimamoto, K. & Ohfune, Y. *Eur. J. Pharmac.* **211,** 195–202 (1992)
18. Shigemoto, R. *et al. Neurosci. Lett.* **153,** 157–160 (1993).
19. Van der Kloot, W. *Prog. Neurobiol.* **36,** 93–130 (1991).
20. Ito, S. & Cherubini, E. *J. Physiol., Lond.* **440,** 67–83 (1991).

ACKNOWLEDGEMENTS. We thank Y. Ohfune and K. Shimamoto for DCG-IV, M. Kuno for advice and A. Uesugi for photographic assistance. Partly supported by research grants from the Ministry of Education, Science and Culture of Japan, the Ministry of Health and Welfare of Japan and the Senri Life Science Foundation.

A *Drosophila* homologue of human Sp1 is a head-specific segmentation gene

Ernst A. Wimmer*†, Herbert Jäckle†, Christine Pfeifle*‡ & Stephen M. Cohen*‡

* Howard Hughes Medical Institute, Department of Cell Biology, Baylor College of Medicine, Houston, Texas 77030, USA
† Abteilung Molekulare Entwicklungsbiologie, Max-Planck-Institut für biophysikalische Chemie, Postfach 2841, D-37018 Göttingen, Germany

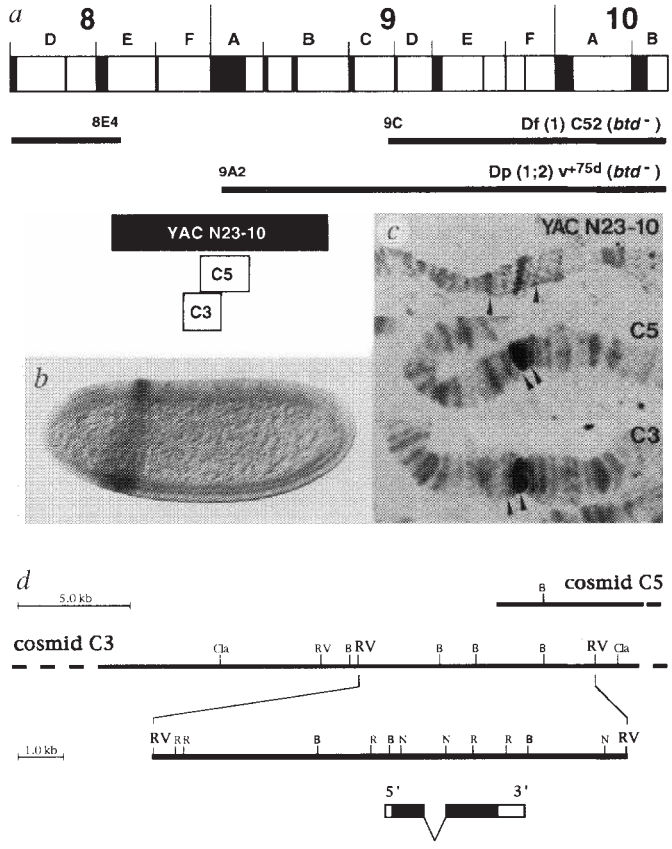
SEGMENTATION in *Drosophila* is based on a cascade of hierarchical gene interactions initiated by maternally deposited morphogens that define the spatially restricted domains of gap gene expression at blastoderm (reviewed in ref. 1). Although segmentation of the embryonic head is morphologically obscured, the repeated patterns of expression of the segment polarity genes reflect the formation of seven head segments[2,3]; two of these depend on the segmentation and homeotic genes used in the trunk, whereas the others form as a result of the activity of the head-specific genes *orthodenticle* (*otd*), *empty spiracles* (*ems*) and *buttonhead* (*btd*). The genes *ems* and *otd* encode homeodomain proteins, suggesting that they may function as transcription factors[4–6]. They are expressed in overlapping stripes in the early embryonic head of *Drosophila*, and their vertebrate homologues, *otx* and *emx*, are expressed in overlapping domains in the anterior central nervous system of the mouse embryo[7,8]. We show here that *btd* is expressed in a stripe covering the head anlagen of the segments affected in *btd* lack-of-function mutants and that *btd* encodes a zinc-finger-type transcription factor with sequence and functional similarity to the prototype mammalian transcription factor Sp1 (ref. 9). When expressed in the spatial pattern of *btd*, a transgene providing Sp1 activity can support development of the mandibular segment in the head of *btd* mutant embryos. A ubiquitous transcription factor from humans can therefore replace an essential component of the genetic circuitry required to specify the development of a particular head segment in the fly.

Positional cloning of the *btd* locus was initiated using a yeast artificial chromosome (YAC) clone[10] (Fig. 1). The identification of the *btd* gene is based on a 10.5-kilobase (kb) transgene which can rescue the *btd* mutant phenotype, and on DNA sequences

‡ Present address: European Molecular Biology Laboratory, Postfach 102209, D-69012 Heidelberg, Germany

of two mutant alleles of the *btd* gene, which reveal alterations in the protein-coding region (Fig. 1). The *btd* transcript is first expressed in a stripe covering the head anlagen of the syncitial blastoderm embryo located between 65 and 77% egg-length (0% is the posterior pole; Fig. 2*a*, *b*), in a region overlapping the domains of *ems* and *otd* expression[4–6]. The *btd* stripe persists through gastrulation (Fig. 2*c*, *d*) and decays during germ band extension. Later in development *btd* is expressed in a complex spatially restricted pattern (not shown).

Sequence of a *btd* cDNA clone revealed that the predicted *btd* protein (Btd) is strikingly similar to known members of the Sp1 family of transcription factors[11] (Fig. 2*e*–*g*). The extensive sequence conservation in the zinc-finger regions of these proteins suggested that Btd might share the DNA-binding sequence specificity of Sp1. To address this question, a fragment of the Btd protein, containing the zinc-fingers, was expressed as a glutathione-*S*-transferase hybrid protein in *Escherichia coli* (GST–Btd). Gel mobility shift assays showed that the GST–Btd



Transgene	Total progeny	FM7 ♂	btdXG; P[btd+]/+ ♂	btdXG/yw ♀	FM7/yw ♀
P[btd+] 467-3	1233	388	35 (9%)	426	384
P[btd+] 467-48	912	234	7 (3%)	351	320
P[btd+] 467-66	635	175	5 (3%)	240	215

FIG. 1 Positional cloning and identification of *btd*. *a*, Cytogenic localization of the *btd* locus. *btd* is included in the deletion Df(1)C52 (8E4-9C), but is excluded from the duplication Dp(1;2)v^{+75d} (9A2-10C). Therefore, the *btd* locus maps to the cytogenetic interval 8E4-9A1. In order to clone the gene, we obtained a YAC clone[10] which covers the cytogenetic interval 8E3-9B12 (see *c*). DNA from YAC N23-10 was used to screen a cosmid library[21]. Cosmids JHD4:39D6 (C3) and JHD5:56D1 (C5) map to the 8F-9A region (see *c*), near the distal breakpoint of Dp(1;2)v^{+75d}. The cosmids were tested for the presence of genes expressed in the early embryo by whole-mount *in situ* hybridization. Cosmids C3 and C5 hybridized to a transcript expressed in a stripe in the head of the syncitial blastoderm embryo covering the anlagen of head segments that are affected in *btd* mutants (see *b*). The head-specific transcription unit was mapped to the region of overlap between the two cosmids. A corresponding cDNA clone was isolated and the extent of the transcription unit was mapped by hybridization to the genomic DNA in cosmid C3 (see *d*). To determine whether this transcription unit corresponds to the *btd* gene, we undertook P-element-mediated transformation to rescue the *btd* mutant phenotype (see *d* and *e*) and cloned and sequenced the DNA of two *btd* mutant alleles. In btd^{XG81}, a nucleotide replacement (G to A transition) leads to the amino-acid replacement C396Y. C396 is the second cysteine residue of the third zinc-finger (Fig. 2*e*). The replacement of C396 by a tyrosine should impair sequence-specific DNA-binding, and is therefore consistent with the lack-of-function phenotype reported for the btd^{XG81} allele[2]. In btd^{XA17}, a nucleotide replacement (C to T transition) introduces a stop codon at amino-acid position 70 (CAA to TAA), which should produce a truncated protein (Fig. 2*e*). *b*, Whole-mount *in situ* hybridization to embryos using cosmid C5 DNA as a probe detects a transcript expressed in a stripe in the head region of the blastoderm stage embryo. Orientation of the embryo is anterior left and dorsal up. *c*, Cytogenetic mapping of the cloned DNA probes by hybridization to polytene chromosomes. Borders of the hybridization signals are indicated by arrowheads. *d*, Structure of the *btd* rescue construct. A 10.5-kb genomic *Eco*RV fragment includes the entire head-specific transcription unit. The structure of the transcription unit is shown below. Coding regions are represented by filled boxes. 5′ and 3′ non-coding sequences are shown as open boxes. B, *Bam*HI; Cla, *Cla*I; N, *Nde*I; R, *Eco*RI; RV, *Eco*RV. *e*, Functional rescue of the *btd* mutant phenotype by the transgene. Females carrying a *btd* mutation (genotype btd^{XG81}/FM7) were crossed to males carrying the transgene on chromosome III; *btd* is located on the X chromosome. As *btd* causes an embryonic lethal phenotype, all males carrying the maternally inherited btd^{XG81} chromosome will die unless they receive a functional copy of the *btd* locus in the transgene. Three independent lines of males carrying the P[w^+(btd^+)] transgene were able to rescue the *btd* mutant phenotype in a fraction of their male progeny. FM7 males can be identified by the following genetic markers: white bar eyes, yellow bodies and forked bristles. The rescued '*btd* mutant' males (carrying the btd^{XG81} chromosome) can be recognized by yellow bodies, red, normally shaped eyes and normal bristles. The number of progeny in the indicated genotypic classes is shown. The proportion of rescued males is shown as per cent of the number of their FM7 brothers (in parentheses). Equivalent results were obtained when the females carried a different *btd* allele (btd^{XA17}; data not shown). No btd^{XG81} or btd^{XA17} male progeny survive when these females are mated to control males that do not carry the P[w^+(btd^+)] transgene (data not shown).

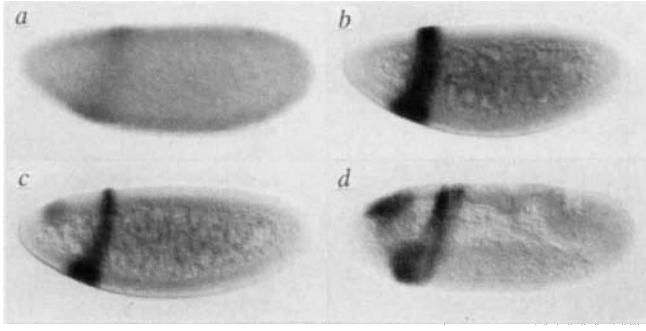
METHODS. DNA from YAC N23-10 was resolved by pulsed-field gel electrophoresis and recovered using NaI/glass powder (Bio 101). *In situ* hybridization to polytene chromosomes and embryos was performed using digoxigenin-labelled DNA probes as described[22]. The signal produced by the whole YAC probe was relatively weak, so to visualize this signal the reaction was developed for 48 h. The substrate was changed after 24 h. A *btd* cDNA clone was isolated from a 0–4 h embryonic cDNA library (gift from N. Brown and F. Kafatos). The extent of the coding region was determined by hybridization of the cDNA clone to genomic DNA, and by DNA sequencing of both genomic and cDNAs. Rescue construct P[w^+(btd^+)] contains a 10.5-kb genomic *Eco*RV fragment cloned into the P-element transformation vector CaSpeR 4 (ref. 23). Transgenic strains were produced by injecting P-element constructs, with pΔ2-3 as helper, into Df(1)w^{67c23}, *y* embryos[24]. Three independently isolated transgenic strains were tested. The transgene insertions in lines 467-3 and 467-48 are balanced over TM3, *Sb*. The transgene insertion in line 467-66 is homozygous on chromosome III. To sequence *btd* mutant alleles, homozygous embryos were selected on the basis of their mutant phenotype. Individual mutant embryos or DNA preparations of 20 mutant embryos have been used for PCR reactions to amplify the *btd* coding region. Several independently obtained clones were sequenced for each allele.

fusion protein binds to an Sp1 consensus binding-site oligonucleotide in a sequence-specific manner (Fig. 3*a*).

The transcriptional activation function of Sp1 is encoded by two serine/threonine-rich domains alternating with two glutamine-rich domains[12,13] (Fig. 2*f*). Btd contains a single Gln-rich domain, and two Ser/Thr-rich domains. In addition, Btd contains an 11-amino-acid motif (Btd box) N-terminal to the zinc-finger domain which is found in 5 out of 6 members of the Sp1-related family[11,14,15] (Fig. 2*g*). Although the function of the Btd box is not known, it may contribute to the ability to activate transcription, because an overlapping 24-amino-acid deletion reduces the Sp1 activation function significantly (see Fig. 5B of ref. 13). In view of the similarity between human Sp1 and Btd, we tested whether Btd could activate transcription from an Sp1 binding site in a co-transfection assay in *Drosophila* Schneider cells. As a target for Btd, we used reporter plasmids carrying zero or two consensus Sp1-binding sites (GC boxes)[16] which were placed upstream of the basal alcohol dehydrogenase promoter in a chloramphenicol acetyltransferase (CAT) reporter construct (Fig. 3*b*). Cells co-transfected with the *btd* expression construct showed a significant stimulation of CAT expression (Fig. 3*c*). We also assessed the ability of Btd to activate transcription from a natural Sp1 target, the SV40 early enhancer[13]. As reported for Sp1, Btd shows a concentration-dependent stimulation of transcription from the SV40/CAT reporter construct (Fig. 3*d*), suggesting that the transcriptional activation properties of Btd are similar to those of Sp1 in Schneider cell co-transfection assays.

To investigate further the functional similarity between Btd and Sp1, we asked whether Sp1 could provide Btd function in the *Drosophila* embryo. We prepared two constructs placing Sp1 coding sequences under control of the *btd* promoter. In one construct, Sp1 complementary DNA was cloned into the 5′ untranslated leader of the *btd* gene. In the second construct, Sp1 was expressed as a fusion protein containing the first 13 amino acids of Btd. The ability of the Sp1 transgenes to rescue the *btd* mutant phenotype was assessed by monitoring the pattern of expression of the segment polarity gene *engrailed* (*en*), and by examination of the embryonic cuticle (Fig. 4). Both transgenes produced equivalent results. In wild type, the maxillary, mandibular, intercalary, antennal and ocular segments can be distinguished by their characteristic stripes of *en* expression[2,3] (Fig. 4*a*, *b*). In *btd* mutant embryos, the mandibular, intercalary and antennal *en* stripes are absent[2] (Fig. 4*c*, *d*). In *btd* mutant embryos that also carry the Sp1 transgene, the mandibular *en* stripe is restored (Fig. 4*e*, *f*), but the antennal and intercalary *en* stripes are still missing. Each of the head segments in which Btd is required develops a characteristic set of cuticular specializations which are missing in *btd* mutant embryos[2] (Fig. 4*g*, *h*). Development of mandibular structures, such as the ventral arms of the head skeleton, is rescued in mutant embryos that carry the Sp1 transgene (Fig. 4*i*).

Sp1 provides only a subset of the functions of the endogenous Btd protein *in vivo*. Btd is required for development of the antennal, intercalary and mandibular segments of the head[2], whereas Sp1 can only support development of mandibular struc-

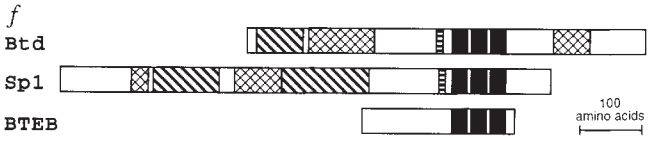


e

Btd

```
MIDAACNYLN PYAQQHQAQQ QQHAQHQQHA QQQQHHLHMQ QAQHHLHLSH QQAQQQHMQH    60
LTQQQQQQQQ QQQQQQQQQQ QQQQPQQQQH DFLSAAALLS APPSLSGSSS GSSSGSSPLY   120
GKPPMKLELP YPQASSTGTA SPNSSIQSAP SSASVSPSIF PSPAQSFASI SASPSTPTTT   180
LAPPTTAAAG ALAGSPTSSS PSSSAASAAA AAAAAAAAAA DLGAAAVASA AYGWNTAYSC   240
LGPARSQFPY AQYASDYYGN AVGMSSSAAW FSHQERLYQP WSSQSYPGFN FDDIAFQTQL   300
QRRSVRCTCP NCTNEMSGLP PIVGPDERGR KQHICHIPGC EELYGKASHL KTHLRWHTGE   360
RPFLCLTCGK RFSRSDELQR HGRTHTNYRP YACPICSKKF SRSDHLSKHK KTHFKDKKSK   420
KVLAAEAKEQ AAAAIKLEKK EKKSGKPLTP PVEFKQEQPD TTPLVNYAPY ANLYQHSTSA   480
GSSVNPPPPP PPLFQQQMTT TTSSAAASFV EQPWSSSSSR AIQPATTSAS SSSSSSASSP   540
AAAVVSAIGS ASSPAASATA LAQHHYAALA MQSESQLAAE YGLTMSGLAS GASQDSSSSC   600
HMKSEYAASY PADFGAGTAS YGYPHPHPHH HNAWAAAYHP HATA                    644
```

f



g

```
Btd    QRRSVRCTCPNCTNEMSGLPPIVGPDERG.RKQHICHIPGCERLYGKASHLKTHLRWHTGERPF
Sp1    T--.EA----Y-KDSEG.....R-SGDP-KK-------Q--GKV---T---RA-----------
SPR-1  L--.-A-S----REGEG.....R-SN-P-KK-------E--GKV---T --RA-----------
SPR-2  L--.-A------KEGGG.....R-TN.L-KK-------GKV---T---RA----S-------
Sp2    K--RMA------KDGEK......RSG-Q-KK-.-V---D-GKTFR-T-L-RA-V-L-------
BTEB   SPEERQDSGSAPSPLSLLHSGVASKGKHASE-R-K-PYS--GKV---S---A-Y-V-------

Btd        LCL..TCGKRFSRSDELQRHGRTHTNYRPYACPICSKKFSRSDHLSKHKKTHFKDKKS
Sp1        M-TWSY-----T--------K----GEKKP---E-P-R-M--------I---.QN--G
SPR-1      I-NWMF-----T--------R----GEKRFE--E---R-M---  ---V---.QN--G
SPR-2      V-NWMY-----T--------R----GEKKFV--E-  -R-M-----A--I---.QN- G
Sp2        V-NWFP-- --T---- ---A----GDKRFE-AQ-Q-R-M---   T--Y---LVT-NL
BTEB       P-TWPD-L-K-- ----T--Y----GEKQFR--L-E-R-M  ---T--ARR-TEFHP
```

FIG. 2 *btd* expression and Btd sequence similarity to Sp1. *a*, *btd* expression is first detectable during nuclear cycle 12/13 as a stripe from 65–77% egg length (100% is the anterior tip of the egg). *b*, Early nuclear cycle 14. *c*, Late nuclear cycle 14. During nuclear cycle 14, the stripe narrows into a wedge shape which is thinner dorsally and wider ventrally. *d*, An additional spot then appears in a more anterior and dorsal position in the head. At the onset of gastrulation, the cephalic furrow that separates the prospective head and trunk regions forms at the posterior edge of the *btd* stripe. *e*, Btd amino-acid sequence deduced from the combined *btd* cDNA and genomic sequences. The zinc-finger region of Btd is underlined, with Cys and His residues shown in bold. The Btd box, an 11-amino-acid motif found in Sp1-related proteins and in Btd (see text, and *g*), is marked. Note that the cysteine residue in position 396 (*) is replaced by tyrosine in the zinc-finger region of the btd^{XG81} allele and that in the btd^{XA17} allele a stop codon is introduced at position 70 (•) (see also legend to Fig. 1*a*). *f*, Schematic representation of the proteins Btd, Sp1 and BTEB. Striped boxes indicate glutamine-rich domains. Cross-hatched boxes indicate serine/threonine-rich domains. Filled boxes indicate the three zinc-fingers (underlined in *e*). The small horizontally striped box adjacent to the finger region represents the novel 11-amino-acid motif (denoted by boxed region in *e*) shared by Btd, Sp1, SPR-1 and SPR-2 (or Sp3) and Sp2. *g*, Comparison of the amino-acid sequences of Btd, Sp1 and Sp1-related proteins[11,14,15] in the zinc-finger region. The 11-amino-acid motif is boxed. Residues identical to Btd are indicated by dashes, and gaps within the sequence by dots. Conserved C_2-H_2 residues of zinc-fingers are in bold. Sequences are taken from ref. 11 (SPR-1 and SPR-2), ref. 14 (BTEB) and ref. 15 (Sp2, Sp3). Note that SPR-2 and Sp3 are identical.

METHODS. The pattern of *btd* expression was determined by whole mount *in situ* hybridization[22] using a *btd* cDNA. The DNA sequence of the *btd* cDNA clone and the flanking genomic DNA was determined. The cDNA clone does not encode the entire *btd* transcription unit. The 5′ portion of the sequence was determined from genomic DNA. The open reading frame of the *btd* transcript extends 516 bp upstream from the 5′ end of the cDNA clone. The predicted Btd open reading frame begins with a consensus *Drosophila* translation initiation sequence[25], just downstream from the second *Bam*HI site in the rescue construct (Fig. 1*d*). We confirmed that this *Bam*HI site is located in the 5′ untranslated leader by preparing a *btd* promoter/*lacZ* fusion construct (described in Fig. 4 legend). Automated DNA sequencing facilities were used at the Institute for Molecular Genetics, Baylor College of Medicine, and at the EMBL.

tures. One explanation may be that Sp1 does not have sufficient activity in *Drosophila* to replace Btd in the antennal and intercalary segments. Alternatively, Btd may interact with the homeodomain proteins Ems and Otd to support development of the anterior head segments[2,17]. Sp1 may therefore activate genes required for development of the mandibular segment that depend only on Btd, but cannot interact with Ems or Otd to activate genes needed in the antennal and intercalary segments.

Sp1 is a member of a larger family of transcription factors and may not itself be the closest vertebrate relative of Btd. Typically, a number of vertebrate genes have been identified as homologues for each of the *Drosophila* segmentation genes. In general it has not been possible to discriminate among the paralogous members of each group on the basis of their sequence similarity to their *Drosophila* counterpart. As nothing is yet known about the pattern of expression of other members of the Sp1 family in the vertebrate embryo, it is possible that one of these genes, or an as-yet unidentified member of the Sp1 family, may have a role in the organization of spatial pattern in the head of the mouse embryo.

Our finding that Sp1 is functionally related to Btd explains the long-standing observation that Sp1 can function as a transcription factor in *Drosophila* tissue culture, although a gene homologous with Sp1 had not been identified in the fly[13,18]. Coactivators essential for the interaction between Sp1 and TATA-binding protein in mammalian systems are also present in the transcription factor TFIID fraction of *Drosophila* nuclear extracts, but are missing from these extracts from yeast, where Sp1 is non-functional (ref. 19, and refs therein). Several of these cofactors, known as TAFs, have been cloned from *Drosophila*; dTAF110 has been shown to interact directly with human Sp1 in *Drosophila* tissue culture cells[20], suggesting that dTAF110 may also serve as a coactivator for Btd in the *Drosophila* embryo. □

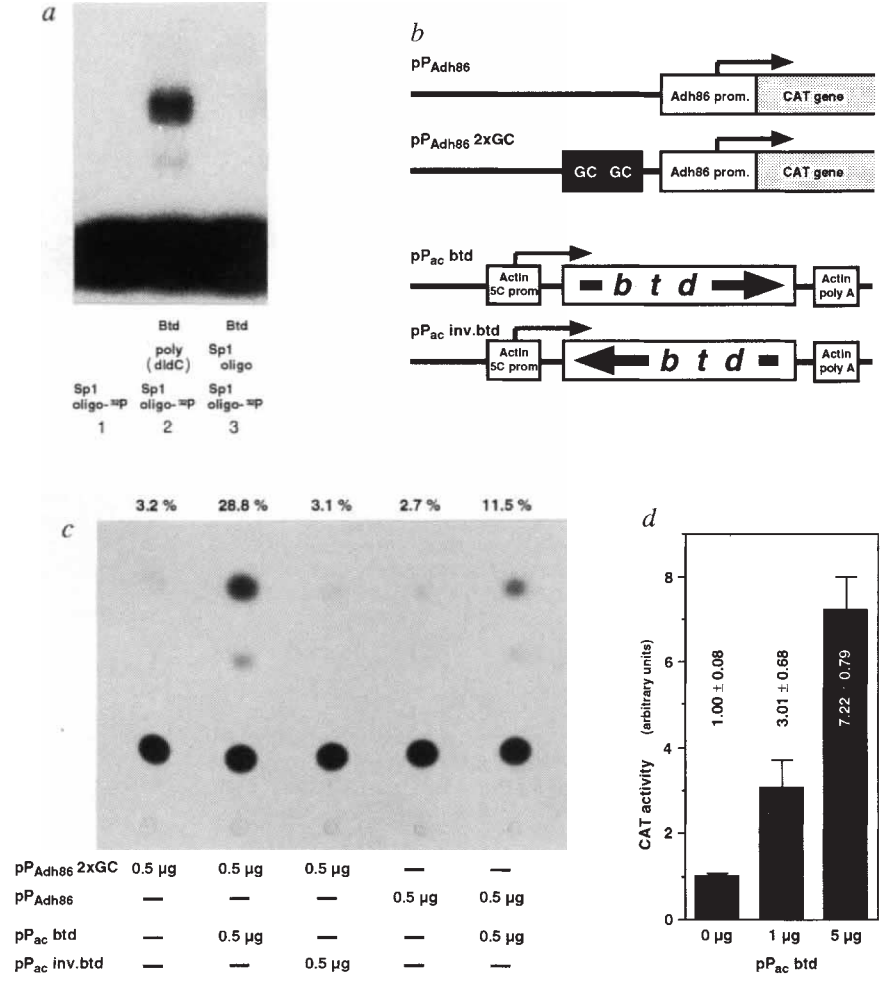


FIG. 3 Sequence-specific DNA binding and transcriptional activation by Btd. *a*, Sequence comparison suggested that Btd would bind to the Sp1 binding site (5′-GGGGCGGGG-3′). Sequence-specific DNA binding of the Btd protein was assessed in a gel mobility shift assay using an Sp1 binding site oligonucleotide, 5′-ATTCGATCGGGGCGGGGCGAGC-3′. Lane 1, end-labelled Sp1 oligonucleotide without added protein; lane 2, with 0.5 µl Btd extract and 1 µg poly(dI·dC); lane 3, with 0.5 µl Btd extract, and 3.5 pmol unlabelled Sp1 oligonucleotide as a specific competitor (1,000-fold molar excess). Btd protein binding to Sp1 oligonucleotide is sequence-specific. *b*, Expression and reporter constructs for transfection of *Drosophila* tissue culture cells. $pP_{Adh\,86}$: control reporter construct in which the basal promoter fragment from the *Drosophila* alcohol dehydrogenase (Adh) gene is linked to the bacterial chloramphenicol acetyltransferase (CAT) gene. $pP_{Adh\,86}2\times GC$: Sp1-sensitive reporter construct containing two additional GC boxes, characterized as Sp1 binding sites upstream of the basal Adh promoter/CAT construct. pP_{Ac}btd: Expression construct in which Btd coding sequences are expressed under control of the constitutive actin 5C promoter. pP_{Ac}-inv.btd: Control expression construct in which Btd coding sequences were inserted in the antisense orientation. *c*, Transcriptional activation by Btd in Schneider cells. *Drosophila* tissue culture cells were co-transfected with a constant amount of DNA (15 µg), including the indicated amounts of the reporter and expression plasmids as described[26]. Percentage indicates fraction of chloramphenicol acetylated in each reaction. The low level of pAdh86 reporter gene activation by Btd is probably due to a weak binding site that is present in the Adh promoter (see Fig. 5 of ref. 27). *d*, Btd activates transcription from a natural Sp1 target, the SV40 enhancer/promoter directing expression of CAT (SV40–BSCAT). Each column represents the mean of three independent transfection experiments; bars indicate standard deviation. The basal level of CAT activity obtained when the reporter construct was transfected alone was assigned the arbitrary value 1.0.

METHODS. Gel mobility shift assays: 50 pg (3.6 fmol) of end-labelled double-stranded Sp1 binding site oligonucleotide was incubated with GST–Btd extract in 25 mM Tris HCl (pH 8.0), 250 mM KCl, 6.25 mM $MgCl_2$, 0.5 mM EDTA, 0.5 mM dithiothreitol (DTT), 10% glycerol for 20 min on ice in a 10-µl reaction volume; 0.5 µl of bacterially expressed GST–Btd fusion protein was used. The glutathione-S-transferase–Btd expression plasmid was prepared by cloning a 615-bp *Nde*I–*Eco*RI fragment of the *btd* cDNA into pGEX-3X. Bacterial extract was prepared as described[9]. The Sp1 consensus binding site oligonucleotide was obtained from Promega. Transfection and CAT assays were carried out as described[26]. Extracts were assayed for CAT activity using ^{14}C-chloramphenicol and the reaction resolved by thin-layer chromatography. $pP_{Adh\,86}$ is described in ref. 27. $pP_{Adh\,86}2\times GC$: Two copies of the Sp1 consensus oligonucleotide were cloned into $pP_{Adh\,86}$; the presence of the insert was confirmed by sequencing the construct. pP_{Ac}btd and pP_{Ac}inv.btd were constructed by cloning a 3.1-kb genomic *Bam*HI fragment containing the Btd coding sequences into the pP_{Ac} expression vector[28]. The SV40–BSCAT reporter contains the SV40 promoter and early enhancer[29].

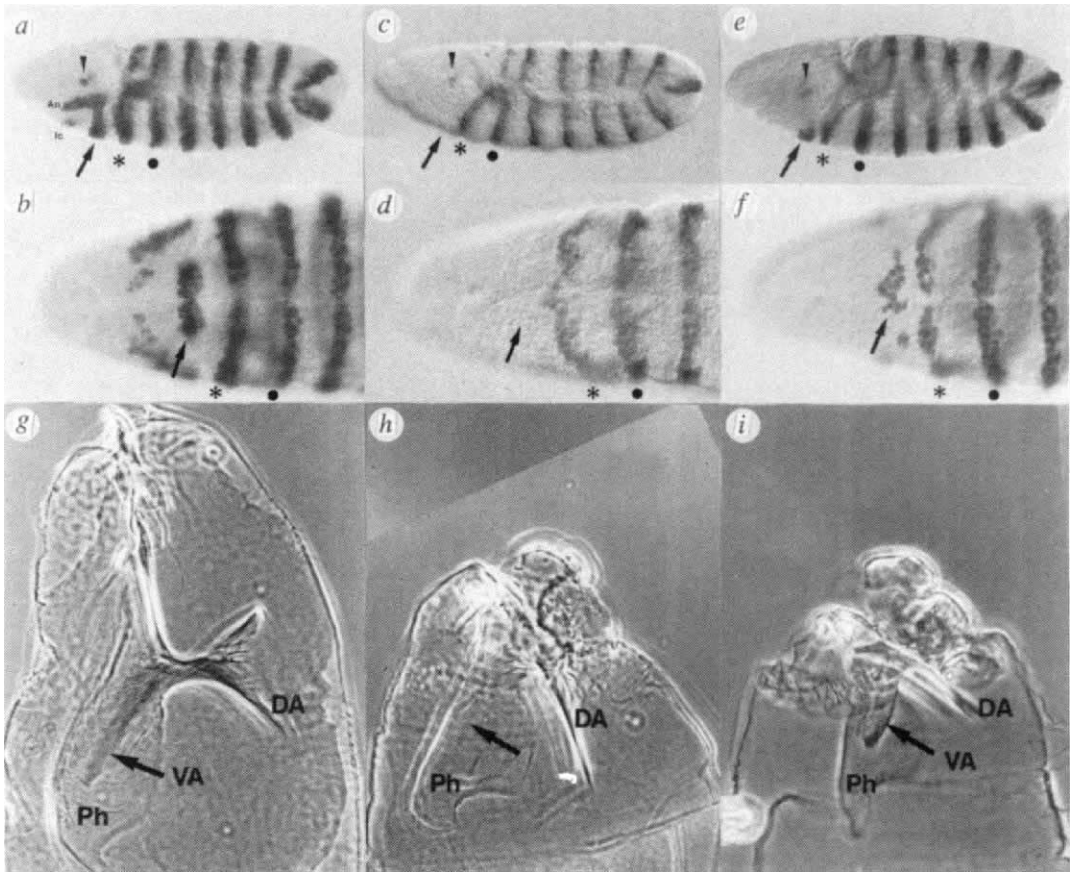


FIG. 4 Sp1 can support development of the mandibular segment. *a–f*, *en* expression reveals the segmental organization of the embryonic head. Orientation of embryos is anterior left. *a* and *b*, Lateral (dorsal up) and ventral views of *en* expression in wild-type embryos. The labial (•), maxillary (*), mandibular (arrow), intercalary (Ic), antennal (An) and ocular (arrowhead) segments can be distinguished by characteristic stripes of *en* expression[2,3]. *c* and *d*, Lateral (dorsal up) and ventral views of *btd* mutant embryos. Labial, maxillary and ocular segments are present. Arrow indicates the absence of the *en* stripe in the mandibular segment. *e* and *f*, Lateral (dorsal up) and ventral views of *btd* mutant embryos carrying the Sp1 transgene. Note that the introduction of the Btd/Sp1 transgene into a *btd* mutant embryo restores expression of *en* in the mandibular segment, but not in the intercalary and antennal segments. *g–i*, The Btd/Sp1 transgene restores cuticular structures missing in the *btd* mutant embryo. *g* and *h*, The cuticle of the head in mature wild type and *btd* mutant embryos (anterior is up and dorsal right). The *btd* mutant embryo lacks structures that derive from the Md, Ic and An segments[2]. The affected mandibular structures include the ventral arms (VA) of the head skeleton, the lateralgräte, and the T-ribs on the floor of the pharynx. Only the VA is indicated. *i*, Development of mandibular structures can be restored by introduction of the Sp1 transgene into *btd* mutant embryos. Note the reappearance of the ventral arms of the head skeleton. Occasionally T-ribs are also produced. As a rescue construct containing the bacterial *lacZ* in place of the Sp1 sequences did not supply any rescuing activity, the rescuing effect of the Sp1 transgene must come from Sp1 activity rather than from 5.2-kb of *Drosophila* sequences present in the transgene (data not shown).

METHODS. Two different Sp1 rescue constructs were prepared. A 5.2-kb genomic fragment from the *Eco*RV site to the second *Bam*HI site of the *btd* gene (Fig. 1*d*) was cloned into a promoterless *lacZ* vector. When provided with a functional promoter and enhancer, this vector is capable of directing *lacZ* mRNA in *Drosophila*. The vector is a derivative of the P-element transformation vector CaSpeR HS43 AUG-β-Gal[23] from which the Hsp43 basal promoter was deleted by partial digestion with *Sal*I. We verified that the promoter construct directed expression of β-galactosidase in the pattern of the endogenous *btd* gene in the embryo, following germline transformation (data not shown). The *lacZ* gene was then replaced with a full-length Sp1 cDNA to make a Btd/Sp1 fusion gene. As the translation initiation context of the Sp1 cDNA is a poor match for the *Drosophila* consensus translation start site[25], a second gene fusion was constructed to link the full-length Sp1 coding sequence to the N terminus of the *btd* coding sequence. The initiator methionine codon of the full-length Sp1 coding sequence[12] encodes amino acid 14 of the fusion gene. We assumed that the presence of 13 additional amino acids of Btd sequence at the N terminus would not adversely affect the function of the Sp1 protein in *Drosophila*. The 5′ untranslated portion of the transcript, including translation initiation sequences are from *btd*. The 3′ end of the transcript, including polyadenylation signals, are from Sp1. Both Btd/Sp1 constructs were expressed in a stripe pattern, indistinguishable from the endogenous *btd* stripe (data not shown), and they produce an equivalent degree of phenotypic rescue following germline transformation.

Received 9 June; accepted 21 September 1993.

1. St Johnston, D. & Nüsslein-Volhard, C. *Cell* **68,** 201–219 (1992).
2. Cohen, S. M. & Jürgens, G. *Nature* **346,** 482–485 (1990).
3. Schmidt-Ott, U. & Technau, G. M. *Development* **116,** 111–125 (1992).
4. Dalton, D., Chadwick, R. & McGinnis, W. *Genes Dev.* **3,** 1940–1956 (1989).
5. Finkelstein, R. & Perrimon, N. *Nature* **346,** 485–488 (1990).
6. Walldorf, U. & Gehring, W. J. *EMBO J.* **11,** 2247–2259 (1992).
7. Simeone, A., Acampora, D., Gulisano, M., Stornaiuolo, A. & Boncinelli, E. *Nature* **358,** 687–690 (1992).
8. Simeone, A. *et al. EMBO J.* **11,** 2541–2550 (1992).
9. Kadonaga, J. T., Carner, K. R., Masiarz, F. R. & Tjian, R. *Cell* **51,** 1079–1090 (1987).
10. Garza, D., Ajioka, J. W., Burke, D. T. & Hartl, D. L. *Science* **246,** 641–646 (1989).
11. Hagen, G., Müller, S., Beato, M. & Suske, G. *Nucleic Acids Res.* **20,** 5519–5525 (1992).
12. Kadonaga, J. T., Courey, A. J., Ladika, J. & Tjian, R. *Science* **242,** 1566–1570 (1988).
13. Courey, A. J. & Tjian, R. *Cell* **55,** 887–898 (1988).
14. Imataka, H. *et al. EMBO J.* **11,** 3663–3671 (1992).
15. Kingsley, C. & Winoto, A. *Molec. cell. Biol.* **12,** 4251–4261 (1992).
16. Dynan, W. S. & Tjian, R. *Cell* **35,** 79–87 (1983).
17. Cohen, S. & Jürgens, G. *Trends Genet.* **7,** 267–272 (1991).
18. Courey, A. J., Holtzman, D. A., Jackson, S. P. & Tjian, R. *Cell* **59,** 827–836 (1989).
19. Dynlacht, B. D., Hoey, T. & Tjian, R. *Cell* **66,** 563–576 (1991).
20. Hoey, T. *et al. Cell* **72,** 247–260 (1993).
21. Hoheisel, J. D., Lennon, G. G., Zehetner, G. & Lehrach, H. *J. molec. Biol.* **220,** 903–914 (1991).
22. Tautz, D. & Pfeifle, C. *Chromosoma* **98,** 81–85 (1989).
23. Thummel, C. & Pirrotta, V. *Dros. Inf. Serv.* **71,** 150 (1992).
24. Rubin, G. M. & Spradling, A. C. *Science* **218,** 348–353 (1982).
25. Cavener, D. *Nucleic Acids Res.* **15,** 1353–1361 (1987).
26. Sauer, F. & Jäckle, H. *Nature* **353,** 563–565 (1991).
27. Heberlein, U. & Tjian, R. *Nature* **331,** 410–415 (1988).
28. Krasnow, M., Saffman, E., Kornfeld, K. & Hogness, D. *Cell* **57,** 1031–1043 (1989).
29. Jones, F. S., Chalepakis, G., Gruss, P. & Edelman, G. M. *Proc. natn. Acad. Sci. U.S.A.* **89,** 2091–2095 (1992).

ACKNOWLEDGEMENTS. We thank F. Sauer for help in setting up the transfection system, J. Hoheisel for cosmid clones, D. Hartl for the YAC clone, G. Chalepakis and P. Gruss for SV40-BSCAT DNA, J. Kadonaga for Sp1 cDNA, P. Lawrence for *en* antibodies, I. Gajewski for preparing photographs, G. Dowe for sequencing, and M. E. McGuffin for help in preparing transgenic fly strains. This work was supported by the Howard Hughes Medical Institute (E.A.W., C.P. and S.M.C.), by the EMBL (S.M.C.), by the Max-Planck Society (E.A.W. and H.J.), and by the Fonds der Chemischen Industrie (H.J.). E.A.W. is a fellow of the Boehringer Ingelheim Fonds.

L-selectin-mediated lymphocyte rolling on MAdCAM-1

E. L. Berg*, L. M. McEvoy, C. Berlin, R. F. Bargatze & E. C. Butcher

Laboratory of Immunology and Vascular Biology, Department of Pathology, Stanford University, Stanford, California 94305, USA, and Center for Molecular Biology and Medicine, Foothill Research Center, Veterans Administration Medical Center, Palo Alto, California 94304, USA

THE L-selectin, a cell surface C-type lectin, directs lymphocyte traffic to lymph nodes[1], and contributes to lymphocyte homing to Peyer's patches[2,3] and to leukocyte interactions with inflamed venules[4,5]. Here we report that the mucosal vascular addressin MAdCAM-1, a mucosal endothelial adhesion molecule with immunoglobulin- and mucin-like domains[6], is a facultative ligand for L-selectin. MAdCAM-1 isolated from mesenteric lymph nodes, but not from cultured endothelioma cells, bears *N*-glycanase-resistant sialic acid-containing carbohydrate which supports adhesion of L-selectin-transfected lymphoid cells under shear. Interacting lymphoid cells display a 'rolling' behaviour similar to the selectin-dependent rolling of neutrophils observed in inflamed venules. MAdCAM-1 is also a ligand for the lymphocyte integrin homing receptor for Peyer's patches, $\alpha 4\beta 7$ (ref. 7), and may be uniquely adapted to support both selectin-mediated lymphocyte rolling and integrin-mediated adhesion and arrest *in vivo*.

High endothelial venule (HEV) ligands for the lymphocyte L-selectin homing receptor display an epitope defined by the peripheral lymph node addressin-specific monoclonal antibody MECA-79: MECA-79 preferentially reacts with L-selectin-binding HEV, inhibits L-selectin-dependent lymphocyte binding to HEV *in vitro* and *in vivo*, and immunoprecipitates HEV molecules that bind lymphocytes and L-selectin transfectants[8,9] or L-selectin-immunoglobulin chimaeric proteins[10,11]. MECA-79 reactivity and L-selectin binding may be conferred by post-translational carbohydrate modifications that are HEV-specific but that can decorate multiple glycoprotein species[8]. When studying MAdCAM-1 isolated from mouse Peyer's patches, we found that it is precipitated not only by anti-MAdCAM-1 antibody MECA-367, but also by MECA-79 (Fig. 1*a*). Serial immunoprecipitation with MECA-79 was unable to preclear MAdCAM-1 from Peyer's patches completely, suggesting that the MECA-79 epitope decorates only a subset of MAdCAM-1 (Fig. 1*b*). MECA-79 also binds in western blots to MAdCAM-1 isolated from mesenteric lymph nodes (MLN) (Fig. 1*c*), confirming that MAdCAM-1 can be directly modified by the MECA-79 epitope.

We also assessed the presence of MAdCAM-1 in MECA-79-isolated addressin from peripheral lymph nodes (PLN) of young mice, whose PLN still retain some of the MAdCAM-1 expression that characterizes PLN-HEV in the early postnatal period. As previously described for human PLN addressin[8], a MECA-79 affinity column binds multiple iodinatible glycoprotein species (Fig. 2). Immunoprecipitated species, which are also recognized prominently by MECA-79 in western blots (not shown), are indicated by arrows (Fig. 2, lane 1). Isolated species include MAdCAM-1, as shown by immunoprecipitation with MECA-367 (Fig. 2, lane 4). The 60K band in MECA-79 immunoprecipitates (double arrows, lane 1) is precleared by anti-MAdCAM-1 antibodies (not shown). These data indicate that MAdCAM-1 can accept the MECA-79 epitope when expressed by HEV. In contrast, MAdCAM-1 isolated from tumour necrosis factor (TNF)-α-stimulated cultured endothelioma cells, or from lymphoid cells transfected with MAdCAM-1 complementary DNA, does not bear the MECA-79 epitope (results not shown, and see below).

The cDNA sequence of MAdCAM-1 predicts a prominent mucin-like domain, a potential site of *O*-glycosylation[6], and three potential *N*-glycosylation sites. Treatment of immunoisolated, iodinated MLN MAdCAM-1 with peptide *N*-glycosidase F (to detect *N*-linked sites) did not alter the relative molecular mass, indicating either that the *N*-glycosylation sites are not used, or that any *N*-linked glycans present are resistant to this enzyme (Fig. 3, lane 2). Neuraminidase treatment increased the migration of MAdCAM-1 in SDS–PAGE (Fig. 3, lane 4). Endo-α-*N*-acetyl-D-galactosaminidase (used to detect *O*-linked sites) had no effect on native (not shown) or neuraminidase-treated MAdCAM-1 (Fig. 3, lane 5). As the activity of this enzyme is largely restricted to a subset of simpler *O*-linked carbohydrates[12], this result is consistent with the presence of complex or modified, sialic acid-bearing *O*-linked carbohydrate. Demonstrable cleavage of the control protein CD44 with all of the enzymes used confirms their activities (Fig. 3, lanes 8–14). It is likely that *O*-linked carbohydrate associated with the mucin-like extracellular domain of MAdCAM-1 represents the site of modification with the MECA-79 epitope.

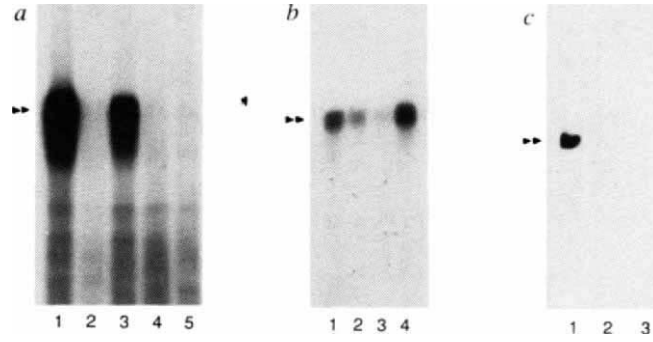


FIG. 1 MECA-79 recognizes MAdCAM-1. *a*, MECA-79 immunoprecipitates purified MAdCAM-1. Immunoisolated MAdCAM-1 from Peyer's patches (PP) was iodinated and precipitated with MECA-367 (lane 1), a rat IgG2a control antibody (lane 2), MECA-79 (lane 3), or rat IgM control antibodies (lanes 4,5). *b*, Only a subset of MAdCAM-1 can be precleared by MECA-79. Immunoisolated, iodinated MAd from PP was precipitated with MECA-79 (lanes 1–3) or MECA-367 (lane 4), before (lane 1) or after preclearing with MECA-79 once (lane 2), twice (lane 3), or three (lane 4) times. The double arrows indicate the M_r of MAdCAM-1 at 58–66K. *c*, MECA-79 reacts with immunoisolated MAdCAM-1 by western blot. Immunoisolated MECA-367 binding material from MLN was applied to an SDS–PAGE gel under non-reducing conditions. After transfer to nitrocellulose, the blot was probed with MECA-79 (lane 1), or two control rat IgM antibodies (lanes 2 and 3). The double arrows indicate the M_r of MAdCAM-1 which is 54–62K under non-reducing conditions[24].

METHODS. *a*, *b*, MAdCAM-1 was prepared from PP stromal lysates of 20 BALB/c mice by immmunoaffinity chromatography with MECA-367 as described[7,24]. A sample from the peak column fraction was then iodinated with 1.8 mCi $Na^{125}I$ (specific activity 15 mCi μg^{-1}) as described[25]. The radiolabelled antigen was immunoprecipitated with the appropriate antibody-coupled Sepharose beads as indicated. For this, samples were diluted with NP-40 wash buffer[7] containing 0.1% haemaglobin (Hb/WB) and the appropriate antibody-coupled beads added. After 30–60 min room temperature incubation, the beads were washed by centrifugation of the beads over 10% and 20% sucrose cushions in Hb/WB, followed by Hb/WB again and then 10 mM Tris–HCl, pH 6.8 containing 0.1% SDS. Material specifically adsorbing to the beads was released by incubation in β-mercaptoethanol-containing solubilization buffer for analysis by SDS–PAGE and subsequent autoradiography. *c*, MLN-MAdCAM-1 was immunoisolated from MLN stromal lysates of 300 mice as previously described[7,21]. Samples of MLN-MAdCAM-1 were run on SDS–PAGE gels, transferred to nitrocellulose paper (NC, BioRad), and western blots done as described[24]. Primary antibodies included 100 $\mu g\ ml^{-1}$ purified MECA-79, or control rat IgM antibodies in 10% horse serum (HS) in RPMI medium. The secondary antibody used was alkaline phosphatase-conjugated rabbit anti-rat IgM (Zymed) in 10% HS in RPMI at 1:200 dilution.

* Present address: Protein Design Labs, Inc., 2375 Garcia Avenue, Mountain View, California 94043, USA.

The presence of the MECA-79 epitope suggested that MAdCAM-1 might be decorated by L-selectin-binding carbohydrates. To test this possibility, glass capillary tubes coated with immunoisolated MLN MAdCAM-1 were infused with mouse L1-2 pre-B cells transfected with human L-selectin. L1-2 cells lack the $\alpha 4\beta 7$ integrin receptor for MAdCAM-1, thus allowing assessment of L-selectin-mediated adhesion in the absence of complicating integrin interactions. As seen in Fig. 4*A*, L1-2 cells transfected with L-selectin (*a*), but not vector (*b*) interact with MLN MAdCAM-1. The interaction is loose, resulting in 'rolling' of transfectants, a behaviour characteristic of selectin-mediated adhesion under flow[5,13]. Rolling is blocked by anti-L-selectin monoclonal antibody Dreg-56 (Fig. 4*A*, *c*) and by preincubation of the coated MAdCAM-1 with neuraminidase (Fig. 4*B*, *a*). Importantly, the L-selectin transfectants displayed no rolling on the MAdCAM-1 isolated from MAdCAM-1-transfected lymphoid cells (not shown) or from TNF-α-stimulated cultured bEnd.3 endothelioma cells (Fig. 4*B*, *b*), although the bEnd.3 MAdCAM-1 was as potent a substrate or slightly better than MLN MAdCAM-1 for $\alpha 4\beta 7$-restricted interactions as assessed by adhesion of TK-1 cells, an $\alpha 4\beta 7$-positive mucosal HEV-binding cell line[7,14] (Fig. 4 legend). We conclude that MAdCAM-1 isolated from lymph nodes supports neuraminidase-sensitive L-selectin-dependent lymphocyte rolling.

In conclusion, MAdCAM-1 expressed by HEV can be modified by carbohydrates comprising the binding sites for the lymphocyte homing receptor L-selectin. Modified MAdCAM-1 represents the first identified high endothelial cell molecule shown to support L-selectin-dependent lymphocyte rolling. Like the independently identified HEV-associated L-selectin ligands GlyCAM-1 (a secreted molecule)[15,16] and CD34[11,17], MAdCAM-1 features a prominent mucin-like domain that may contribute to efficient display of carbohydrate determinants through enhanced valency. MAdCAM-1 is well positioned to contribute to L-selectin-mediated lymphocyte interactions in MLN and Peyer's patches, whose HEV coordinately express high levels of MAdCAM-1 with high (MLN) or low but significant (Peyer's patches) MECA-79 and L-selectin-immunoglobulin reactivity[2,3,18]. It is unlikely to play a significant role as an L-selectin ligand in adult non-mucosal lymph nodes, in which its expression is downregulated, or in lamina propria venules which display MAdCAM-1 but no detectable MECA-79 or L-selectin immunoglobulin binding activity. MAdCAM-1 also binds lymphocyte $\alpha 4\beta 7$ integrin[7], an adhesion pathway that confers specificity to PP homing[19,20] and dominates MAdCAM-1 binding in non-flow *in vitro* assays[7,21]. MAdCAM-1 is thus a multifunctional molecule capable of serving as a vascular receptor for both the leukocyte selectin and for leukocyte integrins. In this regard, MAdCAM-1 represents an elegant example of the expression of multiple adhesive motifs within a single molecule, motifs that may at least cooperate and potentially synergize in support of lymphocyte-HEV interactions. In the context of recent models of lymphocyte-endothelial cell recognition as a multistep process[22,23], MAdCAM-1 expressed in certain *in vivo* sites may in fact be able to participate in both selectin-mediated primary

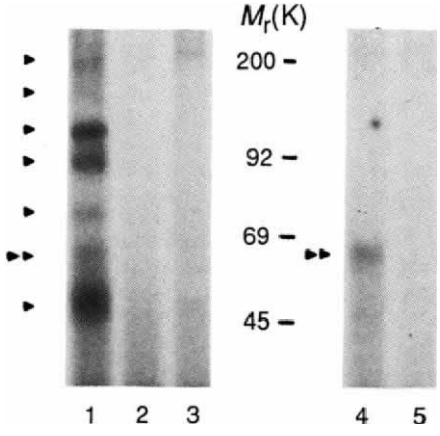


FIG. 2 Anti MAdCAM-1 antibody MECA-367 immunoprecipitates a 60K lymph node antigen that binds to MECA-79. Immunoisolated MECA-79 antigen was iodinated, reprecipitated with MECA-79 (lane 1), control rat IgM antibodies (lanes 2 and 3), MECA-367 (lane 4) or control rat IgG2a antibody (lane 5), and electrophoresed on an SDS-PAGE gel under reducing conditions. The seven major species which are also reactive with MECA-79 in western blots (under both reducing and non-reducing conditions, not shown) are indicated by arrows at 200, 170, 115, 90, 75, ~60 and 50K. The double arrows indicate a 60K species which comigrates with MAdCAM-1. PLN of young BALB/c mice (7 weeks) were used for this experiment because they represent a more concentrated source of MECA-79 antigen, yet still retain some of the MAdCAM-1 expression that characterizes PLN-HEV in the early postnatal period[26]. METHODS. MECA-79 antigens were isolated from PLN stromal lysates of 20 seven-week-old BALB/c mice by immunoaffinity chromatography generally as described previously and above[8,24]. A sample from the peak column fraction was then iodinated and immunoprecipitated as described in Fig. 1.

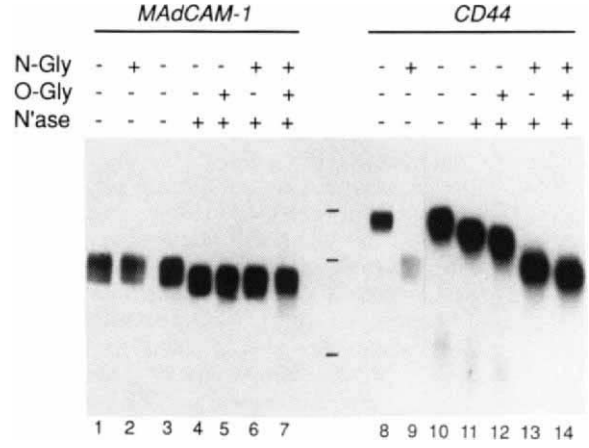
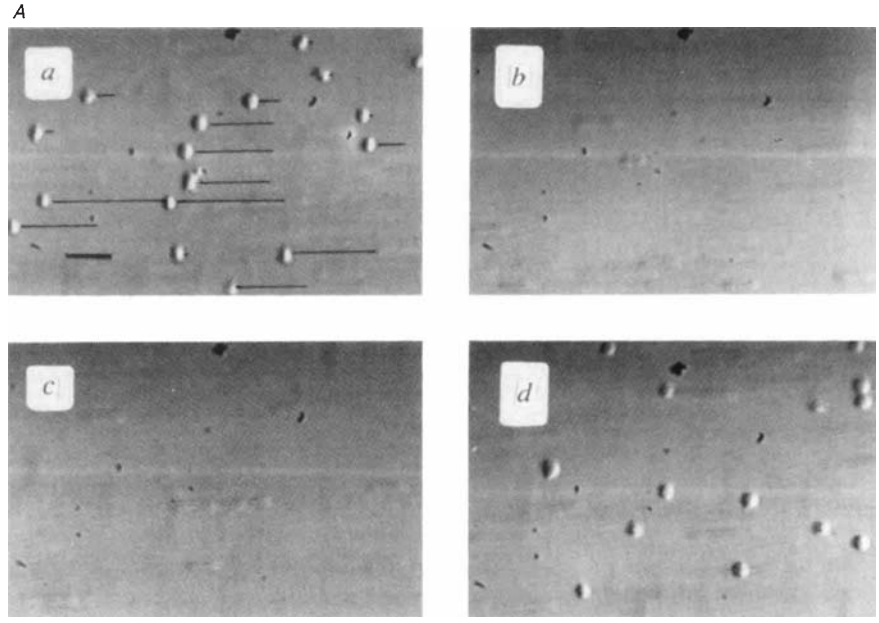


FIG. 3 MAdCAM-1 immunoisolated from MLN contains *N*- and *O*-glycosidase-insensitive sialic acid containing carbohydrate. MAdCAM-1 migration on gels is affected by treatment with neuraminidase but not peptide-*N*-glycosidase F (*N*-Gly) or endo-α-*N*-acetyl-D-galactosaminidase (*O*-Gly) whereas the control glycoprotein CD44, which contains significant amounts of both *N*- and *O*-Gly-sensitive carbohydrate, is affected by all of the enzymes used. Immunoisolated MLN MAdCAM-1, or a control protein, CD44, isolated from human foreskin fibroblast cells[25] was radioiodinated, enzymatically digested, then applied to 8% SDS-PAGE gels under reducing conditions. Autoradiograms prepared from the resulting gels show MAdCAM-1 (lanes 1-7) or CD44 (lanes 8-14) treated with control buffer (lanes 1 and 3; 8 and 10), *N*-Gly (lanes 2 and 9), *Vibrio cholera* neuraminidase (N'ase, lanes 4 and 11), neuraminidase plus *O*-Gly (lanes 5 and 12), neuraminidase plus *N*-Gly (lanes 6 and 13), or all three enzymes (lanes 7 and 14). In experiments not shown, the effect of the neuraminidase treatment on MAdCAM-1 could be blocked by including 10 mM sialyllactose or 5 mM EDTA in the reaction buffer, demonstrating the specificity of the enzyme. METHODS. The MAdCAM-1 used was isolated and iodinated from MLN as described in Fig. 1. Samples were immunoprecipitated with either MECA-367- or anti-CD44-coupled Sepharose beads[25] and then were boiled in 0.15% SDS, 40 mM β-mercaptoethanol before addition of NP-40 for a final concentration of 1.5%. Neuraminidase treatment was done in buffer containing 50 mM NaH_2PO_4, pH 6.0, 5 mM $CaCl_2$ with 2.5 mU *Vibrio cholera* neuraminidase (Boehringer Mannheim) for 6 h at 37°C. Treatment with 0.5 mU endo-α-*N*-acetyl-galactosaminidase (*O*-Gly; Boehringer Mannheim) was done in the same buffer for 16 h at 37°C. For peptide-*N*-glycosidase F (*N*-Gly; Genzyme) digestion, the pH was adjusted to 8.6 before the addition of 0.1 U and incubation for 7-17 h at 37°C. An equal volume of SDS solubilization buffer containing β-mercaptoethanol was added to each sample before SDS-PAGE analysis and subsequent autoradiography.



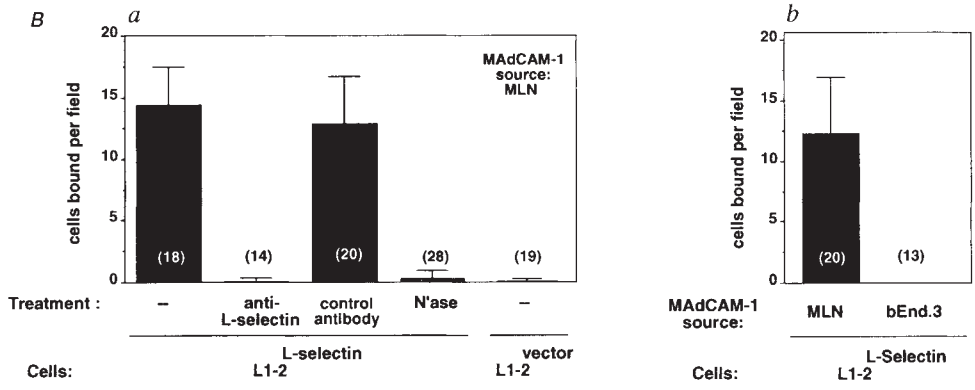


FIG. 4 MAdCAM-1 isolated from mouse MLN supports L-selectin-mediated lymphocyte rolling under shear. *A*, L1–2 cells transfected with L-selectin (L1–2$^{L\text{-}selectin}$) were infused into glass capillary tubes coated with MAdCAM-1 derived from mouse MLN. L1–2$^{L\text{-}selectin}$ cells (*a*) but not control vector transfectants, L1–2^{vector} (*b*) interact with regions of the tube coated with MLN MAdCAM-1. Interaction of L1–2$^{L\text{-}selectin}$ cells with MLN MAdCAM-1 is blocked in the presence of the anti-L-selectin antibody Dreg-56 (ref. 27) (*c*), but not isotype control antibody (*d*). *B*, L-selectin-mediated lymphocyte interactions with MAdCAM-1-coated glass tubes under shear stress. *a*, L-selectin transfectant rolling on MLN MAdCAM-1 is blocked by treatment with the anti-L-selectin antibody Dreg-56 and also by treatment of MLN MAdCAM-1 with neuraminidase. *b*, L-selectin transfectants interact with MAdCAM-1 isolated from MLN but not from TNF-α-stimulated bEnd.3 endothelioma cells. L1–2$^{L\text{-}selectin}$ or L1–2^{vector} transfectants were infused into glass capillary tubes coated with MAdCAM-1. As described, each experiment was recorded on video tape. After a 1-min period to establish equilibrium, the number of cells interacting with the tube in individual frames 4–5 s apart, over the next 90 s, were counted and the average number of cells and the standard deviation obtained is reported. For *b*, both tubes were initiated simultaneously, thus the results obtained for each data point are derived from two 30-s periods of observation over a total of 2 min. The data shown are from individual experiments but were repeated 3–5 times with similar results. The number of frames averaged for each data point is shown in parentheses.

METHODS. MAdCAM-1 was immunoisolated either from MLN stromal lysates from 300 ICR mice or 10 T-175 flasks of the hiMAd-4 subline of the bEnd.3 endothelioma cell line, grown and stimulated for 18 h with 50 U ml^{-1} TNF-α as described[6]. Isolation followed previously described methods[7,8,21]. The bEnd.3 and MLN MAdCAM-1 preparations used were normalized by immunoreactivity with MECA-367 by ELISA. In conventional, non-flow adhesion assays as described by Berlin *et al.*[7], the bEnd.3 MAdCAM-1 was slightly more potent than MLN MAdCAM-1 in mediating $\alpha4\beta7$-restricted adhesion as assessed by the $\alpha4\beta7^{+}$L-selectin^{-} mucosal HEV binding cell line TK-1[7] (the number of cells binding to MLN MAdCAM-1 versus bEnd.3 MAdCAM-1 was 32.4 ± 13.7 s.e.m., $n = 7$ fields, versus 40.1 ± 12 s.e.m., $n = 7$ fields). To examine the ability of MAdCAM-1 to support L-selectin-dependent adhesion under conditions of shear, samples of immunoisolated material were absorbed onto the inside surface of one end of 4-inch-long glass capillary tubes (N51A glass, 1.415 mm internal diameter; Drummond Scientific, Broomall, PA) by dilution of the material in calcium and magnesium-containing Hank's Balanced Salt Solution (HBSS; 1:15). After overnight incubation, the tubes were washed and blocked with 5% normal bovine serum, 20 mM HEPES, pH 7.3 in HBSS for 30 min at room temperature. The development and growth of mouse L1–2 pre-B cell lines transfected with human L-selectin cDNA (L1–2$^{L\text{-}selectin}$) or vector cDNA (L1–2^{vector}) have been previously described[8,28]. Transfectants were grown to 2×10^6 cells per ml and then infused at room temperature in the capillary tube at a flow rate of 2.6 ml min^{-1} with a syringe pump (Sage Instruments, Model 351) fitted with a 12-ml syringe. The coated portion of the tube was ~4 cm from the tube connected to the syringe pump to allow the establishment of laminar flow. This flow rate in this diameter tube corresponds to a wall shear rate of 71 s^{-1} and, assuming a viscosity of the serum-containing medium at room temperature of ~0.015 poise, this corresponds to a wall shear stress of 1.1 dynes cm^{-1} (refs 13, 29) and allows ~4 min continuous observation per sample. The conditions used were not intended to represent physiological conditions which must include among many other parameters, the rheological and anatomical considerations as in ref. 30, which describes the path of lymphocytes through HEV as tortuous and 'pinball-like'. This method does allow the study of adhesive interactions under conditions of defined flow. In some experiments, monoclonal antibodies were added (20 μg ml^{-1}) to the cells 5 min before the assay. The interaction of cells with the tube surface was viewed with a Nikon inverted microscope and the entire experiment recorded with a video monitor. At the flow rate indicated, cells interacting with the tube surface are easily distinguished from non-interacting cells. To quantify the interaction, after an initial 60-s period in which equilibrium was established, three 150 power fields (400×500 μm^2) were observed for 30 s each (for a total observation period of 90 s. On review of the experiments, the video tape was paused every 4–7 s and the number of cells interacting with the tube wall could be counted. In some experiments, the addressin-coated tubes were treated with neuraminidase (*Vibrio cholera* neuraminidase, Calbiochem) 5 mU ml^{-1} in HBSS for 10 min before washing with NBS-containing medium and the addition of cells. All interacting cells were actively rolling, although a few cells (*A*,*a*) would occasionally pause for 2–3 s before moving onward. The mean rolling velocity of L-selectin transfectants on MLN-derived MAdCAM-1 was determined in one experiment and shown to be 141 ± 68 μm s^{-1} s.d. ($n = 20$ cells).

adhesion and rolling under conditions of physiological shear, and in integrin-mediated binding interactions as well, including activation-dependent sticking and arrest. It will be important to assay directly, through *in situ* analysis, the involvement of MAdCAM-1 and of other putative HEV ligands in L-selectin-dependent lymphocyte–HEV interactions *in vivo*. □

Received 4 June; accepted 14 October 1993.

1. Gallatin, W. M., Weissman, I. L. & Butcher, E. C. *Nature* **304,** 30–34 (1983).
2. Hamann, A., Jablonski-Westrich, D., Jonas, P. & Thiele, H.-G. *Eur. J. Immun.* **21,** 2925–2929 (1991).
3. Bargatze, R. F., Streeter, P. R. & Butcher, E. C. *J. Cell Biochem.* **42,** 219–227 (1990).
4. Lewinsohn, D. M., Bargatze, R. F. & Butcher, E. C. *J. Immun.* **138,** 4313–4321 (1987).
5. von Andrian, U. H. *et al. Am. J. Physiol.* **263,** 1034–1044 (1992).
6. Briskin, M. J., McEvoy, L. M. & Butcher, E. C. *Nature* **363,** 461–464 (1993).
7. Berlin, C. *et al. Cell* **74,** 185–195 (1993).
8. Berg, E. L., Robinson, M. K., Warnock, R. A. & Butcher, E. C. *J. Cell Biol.* **114,** 343–349 (1991).
9. Streeter, P. R., Rouse, B. T. N. & Butcher, E. C. *J. Cell Biol.* **107,** 1853–1862 (1988).
10. Imai, Y., Singer, M. S., Fennie, C., Lasky, L. A. & Rosen, S. *J. Cell Biol.* **113,** 1213–1221 (1991).
11. Baumhueter, S. *et al. Science* **262,** 436–438 (1993).
12. Umemoto, J., Bhavanandan, V. P. & Davidson, E. A. *J. biol. Chem.* **252,** 8609–8614 (1977).
13. Lawrence, M. B. & Springer, T. A. *Cell* **65,** 859–873 (1991).
14. Butcher, E. C., Scollay, R. G. & Weissman, I. L. *Eur. J. Immun.* **10,** 556–561 (1980).
15. Lasky, L. A. *et al. Cell* **69,** 927–938 (1992).
16. Brustein, M., Kraal, G., Mebius, R. E. & Watson, S. R. *J. exp. Med.* **176,** 1415–1419 (1992).
17. Simmons, D. L., Satterthwaite, A. B., Tenen, D. G. & Seed, B. *J. Immun.* **148,** 267–271 (1992).
18. Watson, S. *et al. J. Cell Biol.* **110,** 2221–2229 (1990).
19. Holtzmann, B., McIntyre, B. W. & Weissman, I. L. *Cell* **56,** 37–46 (1989).
20. Issekutz, T. B. *J. Immun.* **147,** 4178–4184 (1991).
21. Nakache, M., Berg, E. L., Streeter, P. R. & Butcher, E. C. *Nature* **304,** 32–36 (1988).
22. Butcher, E. C. *Cell* **67,** 1033–1036 (1991).
23. Shimizu, Y., Newman, W., Tanaka, Y. & Shaw, S. *Immun. Today* **13,** 106–110 (1992).
24. Streeter, P. S., Berg, E. L., Rouse, B. T. N., Bargatze, R. F. & Butcher, E. C. *Nature* **331,** 41–46 (1988).
25. Picker, L. J., Nakache, M. & Butcher, E. C. *J. Cell Biol.* **109,** 927–937 (1989).
26. Berg, E. L., Picker, L. J., Robinson, M. K., Streeter, P. R. & Butcher, E. C. in *Cellular and Molecular Mechanisms of Inflammation* Vol. 2 (eds Cochrane, C. G. & Gimbrone, M. A.) 111–129 (Academic, New York, 1991).
27. Kishimoto, T. K., Jutila, M. A. & Butcher, E. C. *Proc. natn. Acad. Sci. U.S.A.* **87,** 2244–2248 (1990).
28. Berg, E. L., Magnani, J., Warnock, R. A., Robinson, M. A. & Butcher, E. C. *Biochem. biophys. Res. Commun.* **184,** 1048–1055 (1992).
29. Perry, M. A. & Granger, D. N. *J. clin. Invest.* **87,** 1798–1804 (1991).
30. Bjerknes, M., Cheng, H. & Ottaway, C. A. *Science* **231,** 402–405 (1986).

ACKNOWLEDGEMENTS. We thank M. Briskin, M. Jutila, S. Michie, P. Streeter, L. Picker and U. von Andrian for discussions and critical review of this manuscript; L. Picker for providing purified CD44 and technical assistance; and S. Grossman for administrative assistance. E.L.B. was a Special Fellow of the Leukemia Society of America. L.M.M. is a fellow of the American Heart Association, California Affiliate and the Multiple Sclerosis Society. C.B. was supported by a fellowship of a special programme to promote epidemiology and rheumatology from the Duetscher Akademischer Austauschdienst, Germany. These studies were supported by the NIH and by an Award from the Veterans Administration.

A first-generation physical map of the human genome

D. Cohen*†, I. Chumakov*† & J. Weissenbach†‡

* Fondation Jean Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France
† Genethon, 1 rue de l'Internationale, BP 59, 91002 Evry Cedex, France
‡ CNRS URA 1445, Institut Pasteur, 25 rue du Dr Roux, 75014 Paris, France

SETS of ordered overlapping cloned genomic DNA fragments that span each of the human chromosomes are urgently needed for identification of human disease genes. Such a physical map also provides unique material to study the structure and function of the genome. We have therefore exhaustively analysed the CEPH yeast artificial chromosome (YAC) library, which contains 33,000 clones, whose insert size was individually determined. These YACs have an average length of 0.9 megabases and cover the equivalent of 10 haploid genomes. Several mapping techniques were combined to provide multiple sources of structural information for most of these clones. Finally, the library was screened with more than 2,000 genetic markers quasiuniformly distributed over 90% of the genome. These results should allow the scientific community to construct detailed maps of all human chromosomes. Moreover, we propose a data analysis strategy that produces a first-generation integrated map covering most of the human genome.

Several different methods have been used to construct genomic physical maps. All these techniques establish overlaps between clones that allow the reconstitution of the original genomic order, but each method has its limitations. We therefore decided to combine four major techniques to generate structural and positional data from a large number of YAC clones.

The restriction fragment pattern of individual clones has been used successfully in the past to establish maps of *Escherichia coli*[1], yeast[2] and *Caenorhabditis elegans*[3]. We produced similar fingerprints for all 33,000 YACs of our library by detection of medium-repeat sequences (THE and L1) containing fragments generated by three enzymes, as described[4].

Single-copy landmark screening also established overlaps[5], and limited application of this approach has already provided maps for two of the smallest human chromosomes[6,7]. If these landmarks are polymorphic sequence-tagged sites (STSs)[7], they can be genetically ordered using meiotic recombination and assigned to specific chromosomes. We used 2,100 polymorphic genetically mapped STSs (ref. 9; and J. W. *et al.*, submitted) to screen partially or totally the same 33,000 YACs. We have obtained an average of 5.52 YACs that are positive for a selected subset of 1,068 STSs, and 2.05 clones for another 714. This allowed us to determine the STS content of 6,580 clones (20% of the total library).

We also used another approach for rapid and extensive single-copy landmark screening based on the use of YACs as hybridization probes. We produced specific sequences for each clone by polymerase chain reaction (PCR) amplification between the ubiquitous Alu repeats[10–12]. We derived PCR products for a subset of 25,000 YACs with a mean size of 1.1 megabases (Mb), representing 9 genome equivalents. After rational pooling[7], these products were spotted at high density onto membranes together with inter-Alu PCR products from a monochromosomal somatic cell hybrid mapping panel[13]. These membranes were successfully hybridized to 5,332 inter-Alu PCR probes derived by amplification of individual YACs representing at least 2 genome equivalents. We also preferentially used polymorphic STSs containing YACs (2100), together with clones from chromosome-specific sublibraries derived for each human chromosome[12]. On average, for each YAC probe, 10 YACs were detected. In this way we established homology relations among clones in a set of 20,750 YACs, that included both YAC probes and their targets. This number corresponds to 83% of the library screened. We were able to assign simultaneously 4,373 YAC probes out of 5,332 YAC probes, to human chromosomes and deduce possible chromosomal locations for more than 15,000 target YACs.

Finally, about 500 YACs containing genetically mapped polymorphic STSs (one every 7.4 centimorgans) were positioned on metaphase chromosomes using FISH[14]. This allowed the integration of genetic, physical and cytogenic maps. All these data, together with clone sizes, are available to the scientific community by anonymous FTP (File Transfer Protocol) from ceph-genethon-map.genethon.fr or by electronic mail requests to ceph-genethon-map@ genethon. genethon.fr.. Specifically, files with STS content of clones, results of Alu PCR hybridizations, L1 and the fingerprint band sizes of individual YACs and relationships between clones, deduced from fingerprint analysis, together with sizes of clones and their FISH localization will be available by FTP. Additional information on the results of screenings will be initially provided through electronic mail contacts. Our library has already been freely distributed since 1992[15].

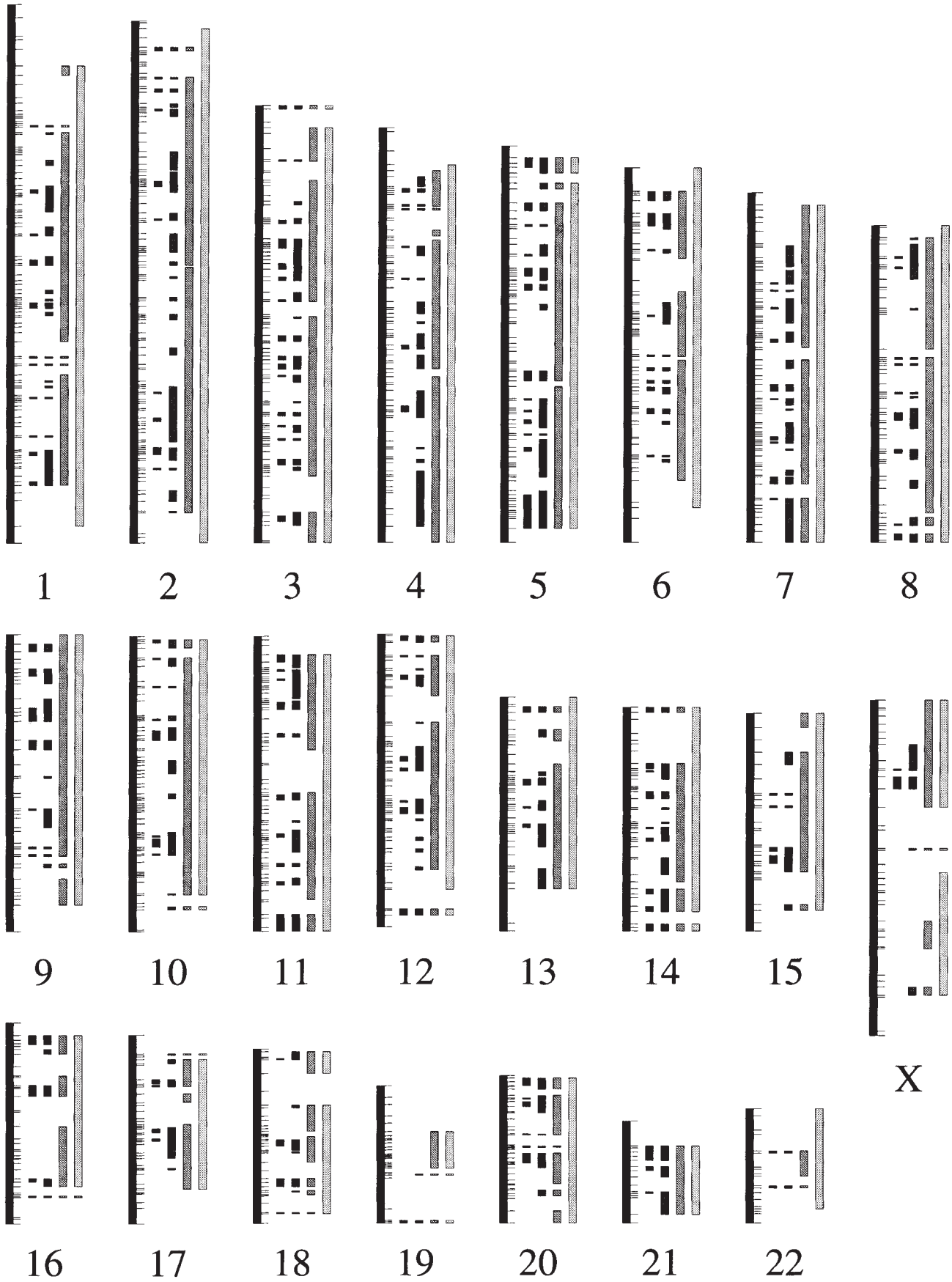


FIG. 1 Schematic representation of genetic distance coverage by tiling path contigs. Genethon 1993 map for each human chromosome as a vertical black bar with horizontal traits representing genetic loci. The distances are proportional to sex-average genetic lengths. Boxes from left to right represent cumulative coverage for levels 1, 3, 5 and 7, respectively.

By combining the data from the experiments described we should ideally be able to reconstruct a continuum of overlapping clones spanning all human chromosomes. The real situation is, however, far from optimal. The main disadvantage of YAC cloning[16–18] is the presence of a large proportion (40–50%) of chimaeric clones, containing artefactually linked segments from noncontiguous portions of the genome. Two types of chimaeric clones could be defined: clones linking segments from different chromosomes, readily detected by chromosomal localization, and syntenic chimaeric clones, which are harder to detect. These artefacts could yield chimaeric continua of clones which contain falsely linked regions of the genome. Another source of artefacts is the presence in the human genome of homologous regions with almost identical sequences which can also result in erroneous contig assembly. In large mapping projects using YAC libraries these problems have been partially solved by cytogenetic methods. In this case, landmarks were reassigned to specific regions with the help of naturally occurring deleted chromosomes[6,7]. Even after such testing, there is however no absolute method to prove that a putative single copy probe detects a single locus.

Here we propose an alternative multilevel mapping approach in which physical map construction is directly based on integration with the genetic map. In this method, continua of overlapping clones are assembled for short, genetically defined adjacent intervals by a chromosome-specific scanning process. The task of our present analysis is to fill with overlapping YACs the genetic intervals corresponding to the current meiotic recombination map of Genethon (J.W. *et al.*, manuscript submitted). This map contains 1,267 genetically resolved loci and covers an estimated 90% of the human genome.

By screening the library with STSs corresponding to genetic loci, we obtained at least one YAC clone for most of them. Taking into account the average 0.9 Mb size for YACs and a quasiuniform distribution of genetic markers, this alone provides physical coverage of 20–30% of the genome with ordered YAC clones. We consider that a gap between neighbouring STSs is closed when a local tiling path can be assembled that starts with a clone containing the first STS and ends with a clone positive for the other one. The maximal number of clones in this tiling path is referred to here as 'level'. Thus, for level 1, only one YAC, positive for both STSs, is used to fill a gap. The results at this level indicate that 11% of the total genetic length is physically covered. During the tiling path assembly we kept only those clones whose chromosomal assignment did not contradict the genetic position of the two neighbouring STSs. In this approach, by neighbouring STSs, we refer to all markers located within a specific genetic interval. Here at level 1, this interval is 10 cM. This allows us to circumvent possible local inversions in the genetic map. Indeed, genetic marker order is only probabilistic and therefore is never 100% certain. In addition, within an interval of this size we can calculate that only 0.12% of the clones will be syntenic chimaeras. The results of this step-by-step gap closure are shown in Fig. 1 and Table 1. The overall physical coverage of the total genetic length at levels 3, 5 and 7 are 30, 70 and 87% respectively. The present map is based on the sex-averaged meiotic map. An example of a tiling path is shown in Fig. 2. The path between *D21S263* and *D21S261* is filled at level 1. The clones 815_f_8 and 968_g_2 establish a level-2 path between *D21S261* and *276za5*, and the five ordered clones establish a level-5 path between *267za5* and *D21S267*. The relations between consecutive clones in these paths are based on Alu PCR hybridization.

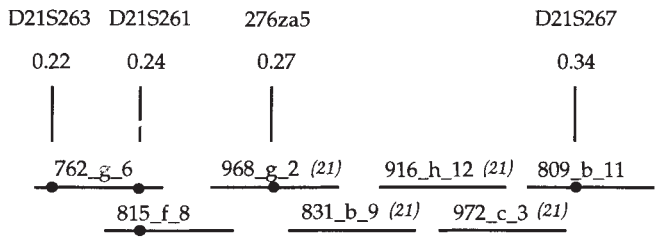


FIG. 2 Tiling paths in a 12-cM region of chromosome 21 containing four STSs. Each STS is represented by a vertical bar. Its positive YACs are marked with a dot. Genetic positions on the Genethon 1993 map are indicated in morgans. When the chromosomal assignment (by Alu PCR hybridization on a panel of somatic cell hybrids) of a YAC is known, it is noted in parentheses.

TABLE 1 Per cent coverage of individual chromosomes at levels 1, 3, 5 and 7

Chromosome	Level 1	Level 3	Level 5	Level 7
1	6.2	21.3	62.2	36.8
2	6.4	36.8	83.9	98.6
3	15.1	30.2	79.7	95.7
4	5.2	39.0	81.4	91.0
5	23.4	46.3	86.1	91.0
6	12.1	21.7	62.6	90.8
7	9.0	50.3	88.9	96.5
8	9.7	32.5	89.6	100
9	14.6	28.0	84.7	91.1
10	10.1	28.1	84.3	87.6
11	18.8	38.8	74.4	93.8
12	11.7	27.4	68.7	88.8
13	7.4	33.6	60.7	82.0
14	18.9	41.7	68.5	94.5
15	7.5	21.7	64.2	90.6
16	12.4	17.1	50.4	76.0
17	6.8	24.2	54.5	69.7
18	10.2	19.5	50.0	75.0
19	2.1	3.1	29.9	29.9
20	17.5	35.0	68.3	98.3
21	20.4	46.9	67.3	67.3
22	4.1	4.1	24.5	87.8
X	4.1	14.5	43.4	69.3

To test our confidence in this approach, we used our previous higher-resolution map deduced for 21q (ref. 7). Some of the YACs in this contig come from our current library. We found a good gap closure already at level-5 between adjacent polymorphic STS, with no contradiction between the two maps, when comparisons could be made. Nevertheless, we are aware that the present map has its limitations. These are due not only to chimaerism and homology problems but also to the non-clonability of certain regions in yeast, the instability of some YACs, the distribution bias of repeats or genetic markers. At present, we do not know the frequency of occurrence of these problems in the human genome. Moreover,, an error rate of the experimental procedure is inevitable. The efficiency of the STS screening was not uniform for all regions of the human genome. Poor coverage of chromosomes 19, 17 and 1p is due mainly to an inability to obtain at least one positive YAC per STS for many loci in these regions. We are currently deriving confirmatory tests, but the pioneer work on *C. elegans*[19] has shown that further international coopration is needed to refine this map. Investigators interested in having access to our present map can obtain detailed tiling path information by requests to electronic mail address ceph-genethon-map@ genethon. genethon.fr.. We hope that new data on overlaps, when integrated in a common database, will increase accuracy and help to exclude false linkages. However, our results can be used whenever genetically defined regions of interest need to be isolated. Access to the structural and positional individual YAC data from this library will certainly help the current mapping effort. □

Received 10 November; accepted 26 November 1993.

1. Kohara, Y., Akyama, K. & Isono, K *Cell* **50,** 495–508 (1987).
2. Olson, M. *et al. Proc. natn. Acad. Sci. Sci. U.S.A.* **83,** 7826–7830 (1986).

3. Coulson, A., Sulston, J., Brenner, S. & Kam, J. *Proc. natn. Acad. Sci. U.S.A.* **83,** 7821–7825 (1986).
4. Bellanné-Chantelot C. *et al. Cell* **70,** 1059–1068 (1992).
5. Barillot, E., Dausset, J. & Cohen, D. *Proc. natn. Acad. Sci. U.S.A.* **88,** 3917–3921 (1991).
6. Foote, S., Vollrath, D., Hilton, A. & Page, D. *Science* **258,** 60–66 (1992).
7. Chumakov, I. *et al. Nature* **359,** 380–386 (1992).
8. Olson, M. V., Hood, L., Cantor, C. R. & Botstein, D. *Science* **245,** 1434–1435 (1989).
9. Weissenbach J. *et al. Nature* **359,** 794–801 (1992).
10. Nelson, D. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **86,** 6686–6690 (1989).
11. de Jong, P. J. *et al. Cytogen. Cell Genet.* **51,** 985 (1989).
12. Chumakov, I. M. *et al. Nature Genet.* **1,** 222–225 (1992).
13. Mullivor, R. A., Greene, A. E., Drwinga, H. L., Toji, L. H. & Kim, C. *Am. J. hum. Genet.* Suppl. **49,** 370 (1991).
14. Ried, T., Baldini, A., Rand, T. C. & Ward, D. C. *Proc. natn. Acad. Sci. U.S.A.* **89,** 1388–1392 (1992).
15. van Ommen, G. J. *Genome Priority Rep.* 1885–1888 (HUGO Bulletin, 1993).
16. Burke, D. T., Carle, G. F. & Olson, M. V. *Science* **236,** 806–812 (1987).
17. Green, E. D., Riethman, H. C., Dutchik, J. E. & Olson, M. V. *Genomics* **11,** 658–659 (1991).
18. Vollrath, D. *et al. Science* **258,** 52–59 (1992).
19. Sulston, J. *et al. Nature Genet.* **356,** 37–41 (1992).

ACKNOWLEDGEMENTS. This work, soon to be published in detail, was accomplished by a multidisciplinary team in which D.C. and I.C. defined the strategy and directed the work. YAC construction, sizing and quality control was done by D.Le Paslier, A. Billault, P. Ougen, M. Saumier, C. Soravito, V. Perrot, F. Dukacz and S. Lecoulant. J.W., C. Sizamis, G. Gyapay and A. Vignal provided unpublished genetic map data and materials and helped with STS content analysis. C. Bellanné-Chantelot conducted fingerprint production with the help of S. Beaufils, I. Georges, F. Glibert, I. Gros, L. C. Susini, V. Morel, C. Geraldes, G. Lucotte, S. Bertrand, M. Sangonard and J. Meny. STS content screening was done by S. Guillou, D. Delporte, G. Guasconi, P. Gervy, F. Gilbert, R. Bahouayila and S. NGuyen. The Alu-PCR hybridization project was realized by I.Le Gall and P. Soularue, H. Bui, J.-L. Sambucy, M. Belova, C. Massart, P. Millasseau, J.-F. Prud'homme and M.-F. De Tand. FISH mapping was conducted by D. Ward, J. Menninger, J. Lieman, T. Desai, A. Banks, P. Tsai, I. P. Bzay-Wazd and, D. Cherif. Seventy-seven chromosome 12 STSs were provided by R. Kucherlapati, K. Montgomery and J. LeBlanc. D. Grausz helped with documentation and communication between the teams. Automation was developed by C. Schmidt and L. Saint-Marthe. G. Vaysseix supervised the computer structure. The computer system was administered by C. Scarpelli, P. Gesnouin and R. Gavrel. Computer support of Alu PCR and STS data production and some of the storage and retrieval systems and analysis were developed by P. Rigault and E. Poullier. Some tiling path algorithms and graphical programs were also developed by P. Rigault (data shown). Other approaches to contig assembly and data treatment were supervised by E. Barillot and G. Vaysseix. Specifically: data capturing of extended structural pattern was done by P. Gesnouin, P. Rigault, E. Barillot, B. Lacroix, J.-J. Codani and E. Poullier, R. Gavrel, S. Bertrand, G. Lucotte, J. Meny and M. Sangouard; fingerprint analysis was done by B. Lacroix, J.-J. Codani and E. Barillot; contig assembly software was developed by E. Barillot and B. Lacroix. Some tiling path algorithms were developed by L. Atlan, M. Tissot, B. Lacroix (data not shown); database development by E. Viara and S. Pook, database management by E. Viara, P. Rodriguez-Tome, X. Begnini, E. Barillot and B. Lacroix. Graphical software was developed by S. Pook and E. Viara. Theoretical calculations and statistical analysis were done by E. Barillot, B. Lacroix and A. Cohen-Akenine. This work was supported by Association Francaise contre les Myopathies, Ministère de la Recherche et l'Enseignement Superieure, Groupement de Recherche et d'Etude des Genomes, la Ligue contre le Cancer, the NIH, and the E.E.C programme DGXII. We acknowledge support at all stages of this project by G. Peirrano, C. Rebollo and all personnel of CEPH and Genethon. This work is dedicated to J. Dausset, B. Barataud and J. Watson.

p21 is a universal inhibitor of cyclin kinases

Yue Xiong*†‡, Gregory J. Hannon*†, Hui Zhang*†, David Casso*†, Ryuji Kobayashi† & David Beach*†§

* Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA
† Cold Spring Harbor Laboratory, 1, Bungtown Road, Cold Spring Harbor, New York 11724, USA

DEREGULATION of cell proliferation is a hallmark of neoplastic transformation. Alteration in growth control pathways must translate into changes in the cell-cycle regulatory machinery, but the mechanism by which this occurs is largely unknown. Compared with normal human fibroblasts, cells transformed with a variety of viral oncoproteins show striking changes in the subunit composition of the cyclin-dependent kinases (CDKs)[1]. In normal cells, CDKs exist predominantly in multiple quaternary complexes, each containing a CDK, cyclin, proliferating cell nuclear antigen and the p21 protein. However, in many transformed cells, proliferating cell nuclear antigen and p21 are lost from these multiprotein enzymes. Here we have investigated the significance of this phenomenon by molecular cloning of p21 and *in vitro* reconstitution of the quaternary cell-cycle kinase complexes. We find that p21 inhibits the activity of each member of the cyclin/CDK family. Furthermore, overexpression of p21 inhibits the proliferation of mammalian cells. Our results indicate that p21 may be a universal inhibitor of cyclin kinases.

In mammalian cells, cell-cycle regulation is effected by the ordered activation of a group of related enzymes known as the cyclin-dependent kinases (CDC2 and CDK2 to CDK5; see ref. 2 for review). The activity of these enzymes is controlled in part by their association with regulatory subunits called cyclins (cyclins A to E; ref. 2). Early studies suggested that multiple binary cyclin/CDK complexes comprised the primary effectors of cell-cycle progression and that further regulation is achieved by both activating and inhibitory phosphorylation of the CDK subunits (reviewed in ref. 3). However, these studies were carried out almost exclusively in transformed cells, and subsequent investigation of normal cells revealed a more complicated situation[4]. In untransformed human fibroblasts, each of the cyclin-dependent kinases exists in part in quaternary complexes consisting of a cyclin, a CDK, proliferating cell nuclear antigen (PCNA) and an uncharacterized subunit of M_r 21K, termed p21 (refs 4, 5). On transformation of fibroblasts by the oncoproteins of various DNA tumour viruses (for example, SV40 T antigen, adenovirus E1A/E1B, HPV-16, E6/E7), the subunit composition of cyclin–CDK complexes undergoes dramatic rearrangement[1]. CDK4 dissociates totally from cyclin D, PCNA and p21, and instead associates exclusively with a polypeptide of M_r 16K (p16). Neither PCNA nor p21 associate with CDC2–cyclin B1. Cyclin A complexes no longer contain p21 but acquire a new 19K subunit. Strikingly, loss of p21 from the CDK complexes also occurs in p53-deficient cells derived from Li–Fraumeni patients that carry no known DNA tumour virus[1]. This indicates that loss of p53 alone might alter the subunit composition of the cell-cycle kinases.

To investigate this possibility, we isolated a complementary DNA encoding the one uncharacterized component of the quaternary complex, as a prerequisite to *in vitro* reconstitution and biochemical analysis. p21 was purified by large-scale anti-cyclin D immunoprecipitation (see Fig. 1 legend). The purified protein was microsequenced and degenerate polymerase chain reaction (PCR) primers were prepared based on the sequences of four peptides. Amplification of cDNA from U118 glioblastoma cells with primers derived from the K10 and K13 peptides (see Fig. 1) generated a 96-base-pair (bp) fragment. This was used as a probe to isolate a full-length cDNA from a library prepared from the U118 cell line, and several positive clones were obtained. One contained an ~600-base insert which included a 164-amino-acid open-reading-frame (see Fig. 1). *In vitro* translation of this cDNA generated a protein which comigrated exactly with the cyclin D-associated p21 from W138 cells (Fig. 2*a*). Partial V8 protease mapping demonstrated that these proteins are identical (Fig. 2*b*).

Cyclin/CDK complexes containing p21 were reconstituted using a baculovirus expression system. Quaternary complex formation was observed in lysates from insect cells carrying viruses that encode p21 and PCNA in combination with those encoding cyclin D and CDK4, cyclin E and CDK2, cyclin A and CDK2 or cyclin B and CDC2 (Fig. 3). These *in vitro* quaternary complexes precisely mirrored our previous data which indicated that p21 and PCNA associate with each of the cyclin kinases *in vivo*[4,5]. Furthermore, p21 alone can bind to each cyclin/CDK pair to form apparent ternary complexes (Fig. 3), suggesting that PCNA is not necessary for the interaction of p21 with cyclins

‡ Present address: University of North Carolina, Chapel Hill, Room 214, Fordham Hall, Chapel Hill, North Carolina 27599, USA.
§ To whom correspondence should be addressed.

```
        10                  30                  50
CCGAAGGTCAGTTCCTTGTGGAGCCGGAGCTGGGCGCGGATTCGCCGAGGCACCGAGGCA

        70                  90                  110
CTCAGAGGAGGCGCCATGTCAGAACCGGCTGGGGATGTCCGTCAGAACCCATGCGGCAGC
                M  S  E  P  A  G  D  V  R  Q  N  P  C  G  S

        130                 150                 170
AAGGCCTGCCGCCGCCTCTTCGGCCCAGTGGACAGCGAGCAGCTGAGCCGCGACTGTGAT
K  A  C  R  R  L  F  G  P  V  D  S  E  Q  L  S  R  D  C  D

        190                 210                 230
GCGCTAATGGCGGGCTGCATCCAGGAGGCCCGTGAGCGATGGAACTTCGACTTTGTCACC
A  L  M  A  G  C  I  Q  E  A  R  E  R  W  N  F  D  F  V  T

        250                 270                 290
GAGACACCACTGGAGGGTGACTTCGCCTGGGAGCGTGTGCGGGGCCTTGGCCTGCCCAAG
E  T  P  L  E  G  D  F  A  W  E  R  V  R  G  L  G  L  P  K

        310                 330                 350
CTCTACCTTCCCACGGGGCCCCGGCGAGGCCGGGATGAGTTGGGAGGAGGCAGGCGGCCT
L  Y  L  P  T  G  P  R  R  G  R  D  E  L  G  G  G  R  R  P
                    K21
        370                 390                 410
GGCACCTCACCTGCTCTGCTGCAGGGGACAGCAGAGGAAGACCATGTGGACCTGTCACTG
G  T  S  P  A  L  L  Q  G  T  A  E  E  D  H  V  D  L  S  L

        430                 450                 470
TCTTGTACCCTTGTGCCTCGCTCAGGGGAGCAGGCTGAAGGGTCCCCAGGTGGACCTGGA
S  C  T  L  V  P  R  S  G  E  Q  A  E  G  S  P  G  G  P  G
                    K10
        490                 510                 530
GACTCTCAGGGTCGAAAACGGCGGCAGACCAGCATGACAGATTTCTACCACTCCAAACGC
D  S  Q  G  R  K  R  R  Q  T  S  M  T  D  F  Y  H  S  K  R
                                    K13
        550                 570                 590
CGGCTGATCTTCTCCAAGAGGAAGCCCTAATCCGCCCACAGGAAGCCTGCAGTCCTGGAA
R  L  I  F  S  K  R  K  P  *
   K14
```

FIG. 1 Sequence of the p21 cDNA. The sequence of the p21 cDNA and the deduced amino-acid sequence of the p21 protein are shown. The positions of the peptides obtained from microsequencing are underlined (see text). Asterisk indicates stop codon.

METHODS. For the purification of p21, WI38 cells (passage 24) were grown to late log phase on 400 15-cm tissue culture dishes in Dulbecco's minimal essential medium supplemented with 10% fetal bovine serum. All subsequent steps were carried out at 4 °C. Cells were rinsed in phosphate-buffered saline (PBS), collected and lysed in NP40 lysis buffer[1]. The lysate was clarified by centrifugation at 5,000g for 10 min, and the supernatant was added to 2.5 mg of affinity-purified cyclin D1 antibodies, covalently crosslinked to protein A–Sepharose beads with dimethyl palmitate (Pierce). Immunoprecipitation was carried out for 4 h, after which the beads were washed four times with lysis buffer. The supernatant was subjected to a second round of affinity chromatography by the addition of fresh anti-cyclin D1–Sepharose. Proteins were released from the beads by boiling in SDS sample buffer and loaded onto a 12% SDS–polyacrylamide gel (acrylamide:bis-acrylamide 37.5:1). The gel was stained with 0.05% Coomassie brilliant Blue G (Sigma) for 15 min, destained and soaked in water for 1 h. The p21 protein (~2 μg) was excised from the gel and subjected to in-gel digestion with *Achromobacter* endoprotease I for 24 h at 30 °C as described previously[9]. Digests were cleared by centrifugation and the supernatant was filtered through a 0.22-μm membrane (Ultrafree-MC, Millipore). Peptides were separated by reverse-phase HPLC (Hewlett Packard 1090) using a Vydac C18 column (2.1 × 250 mm, 5 μm, 300 A) with an anion-exchange pre-column (Brownlee GAX-013, 3.2 × 15 mm). Peptides were eluted with an acetonitrile gradient and their absorbance was monitored at 214, 280, 295 and 550 nm. Amino-acid sequencing was performed on an automated microsequencer (ABI model 470) with on-line HPLC (ABI model 120A) analysis of phenylthiohydantoin-tagged amino acids. Four peptide sequences were obtained as follows: K10, tSLVP?SGEQAEGSPsk; K13, ?RQTSMTDFYHSK; K14, RRLIFSK; K21, LYLPTGPpRSRDEDLGIS. Lower-case letters indicate likely identification and question marks indicate uncertainty. Degenerate oligonucleotide primers were designed based on the three longest peptides, K10, K13 and K21. Primers were synthesized in both directions and codons for serine were split to reduce degeneracy. One or two amino acids were reserved at each end for the verification of any PCR product. Template DNA was prepared from several human cDNA libraries (WI38 and HeLa cell, Stratagene; U118 cell, a gift from M. Wigler) as described[10]. Thirty-five cycles of amplification were performed (50 or 55 °C annealing and 72 °C extension) after which products were resolved on a 1.5% agarose gel. All DNA fragments smaller than 500 bases were excised and inserted into pUC118 for sequence analysis. Amplification with one pair of primers derived from the K10 and K13 peptides (K10 primer GGNGA[A/G]CA[A/G]GCNGGNTCNCC, encoding GEQAEGSP; K13 primer TTNCA[A/G]TG[A/G]TA[A/G]AA[A/G]TCNGTCATNCANGT[C/T]TG, encoding QTSMTDFYHSK) gave rise to a specific 96-bp PCR fragment that encoded 13 amino acids in addition to those encoded by the PCR primers. This sequence encoded KRR preceding the K13 primer. This was consistent with the sequence K?R which was predicted to precede the sequence encoded by the K13 primer. The PCR fragment was excised from pUC118 and used as a probe for screening a human U118 cDNA library. Several positive clones were obtained, and 10 were chosen for subsequent analysis. One of these contained an ~600-base insert which was shown to carry the entire p21-coding region.

and CDKs. We cannot, however, formally exclude the possible presence of insect cell PCNA in the p21/CDK/cyclin complexes.

We used the same baculovirus reconstitution system to assess the effect of p21 on the enzymatic activity of the cyclin kinases. Anti-CDK2 immunoprecipitates were prepared from lysates of ^{35}S-labelled insect cells co-infected with viruses encoding CDK2 and cyclin A. The precipitates contained cyclin A/CDK2 and displayed substantial histone H1 kinase activity (Fig. 4*a–d*). Addition of increasing amounts of a lysate containing p21 resulted in progressive formation of cyclin A/CDK2/p21 ternary complexes (Fig. 4*a*). As p21 neared stoichiometry, CDK2 kinase activity was abruptly inhibited (Fig. 4*b*). Addition of PCNA, in an otherwise identical experiment, resulted in the formation of quaternary complexes (Fig. 4*c*) but had no effect on either histone kinase activity or its inhibition by p21 (Fig. 4*d*). Similar inhibition by p21 was observed with cyclin B/CDC2 and cyclin E/CDK2 kinases (our unpublished results).

Cyclin/CDK4 kinase differs from the other cyclin kinases in its inability to utilize histone H1 as a substrate. To date, the only substrates known for cyclin D/CDK4 kinases are the members of the retinoblastoma (Rb) family of 'pocket' proteins[6]. We therefore tested the effect of p21 on the ability of cyclin D/CDK4 to phosphorylate Rb. As previously described, insect cell lysates containing cyclin D or CDK4 alone showed little activity towards glutathione *S*-transferase (GST) fused to Rb.

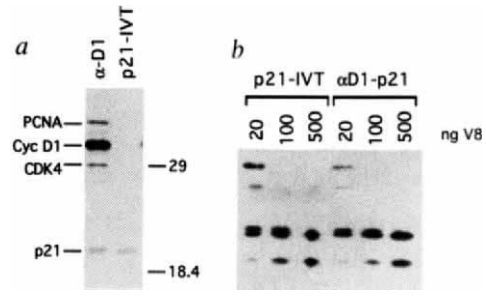


FIG. 2 *a*, Cyclin D1-associated proteins were immunoprecipitated from ^{35}S-methionine-labelled WI38 cells using an anti-cyclin D1 antiserum (α-D1, left lane) and electrophoresed in parallel with *in vitro*-translated, ^{35}S-methionine-labelled putative p21 protein (p21-IVT, right lane). For reference, the positions of labelled protein markers are shown. *b*, *In-vitro*-translated putative p21 and cyclin D1-associated p21 were electrophoretically purified from a gel similar to that shown in *a*. These polypeptides were then subjected to partial V8 protease digestion using the indicated amounts of enzyme. Resulting peptide fragments were fractionated on a 17.5% polyacrylamide gel. Cell culture, immunoprecipitation and partial V8 protease digestion were performed exactly as described previously[4].

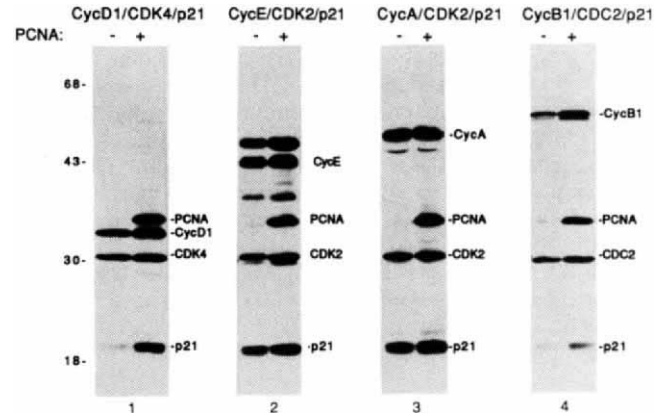


FIG. 3 Reconstitution of p21/PCNA/cyclin/CDK quaternary complexes in baculovirus-infected cells. Exponentially growing Sf9 cells were co-infected with baculoviruses encoding p21 and the indicated cyclins and CDKs in the presence (+) or absence (−) of a baculovirus directing the expression of PCNA. At 40 h post-infection, cells were labelled with ^{35}S-methionine for 3 h and lysed in NP40 lysis buffer. The protein complexes were immunoprecipitated with anti-CDK4 (panel 1), anti-CDK2 (panels 2, 3) or anti-CDC2 (panel 4) antibodies and analysed by gel electrophoresis. The positions of protein M_r standards are indicated.

METHODS. Insect Sf9 cells were maintained in Grace's medium supplemented with 10% heat-inactivated fetal bovine serum, lactalbumin hydrolysate and yeastolate ultrafiltrate at 27 °C. Recombinant baculoviruses were constructed using the Baculo-Gold transfection system (Pharmingen) according to the manufacturer's instructions. CDK2, CDC2, cyclin A, cyclin D1 and cyclin E baculoviruses were gifts from D. Morgan, C. Sherr or S. Gruenwald. For the production of recombinant proteins, Sf9 cells were infected with multiplicities of infection above 10. After 36–48 h, the cells were labelled with 100 μCi ml^{-1} of ^{35}S-methionine (NEN) for 3 h. Cell labelling, cell lysis and immunoprecipitation were carried out as described previously[11].

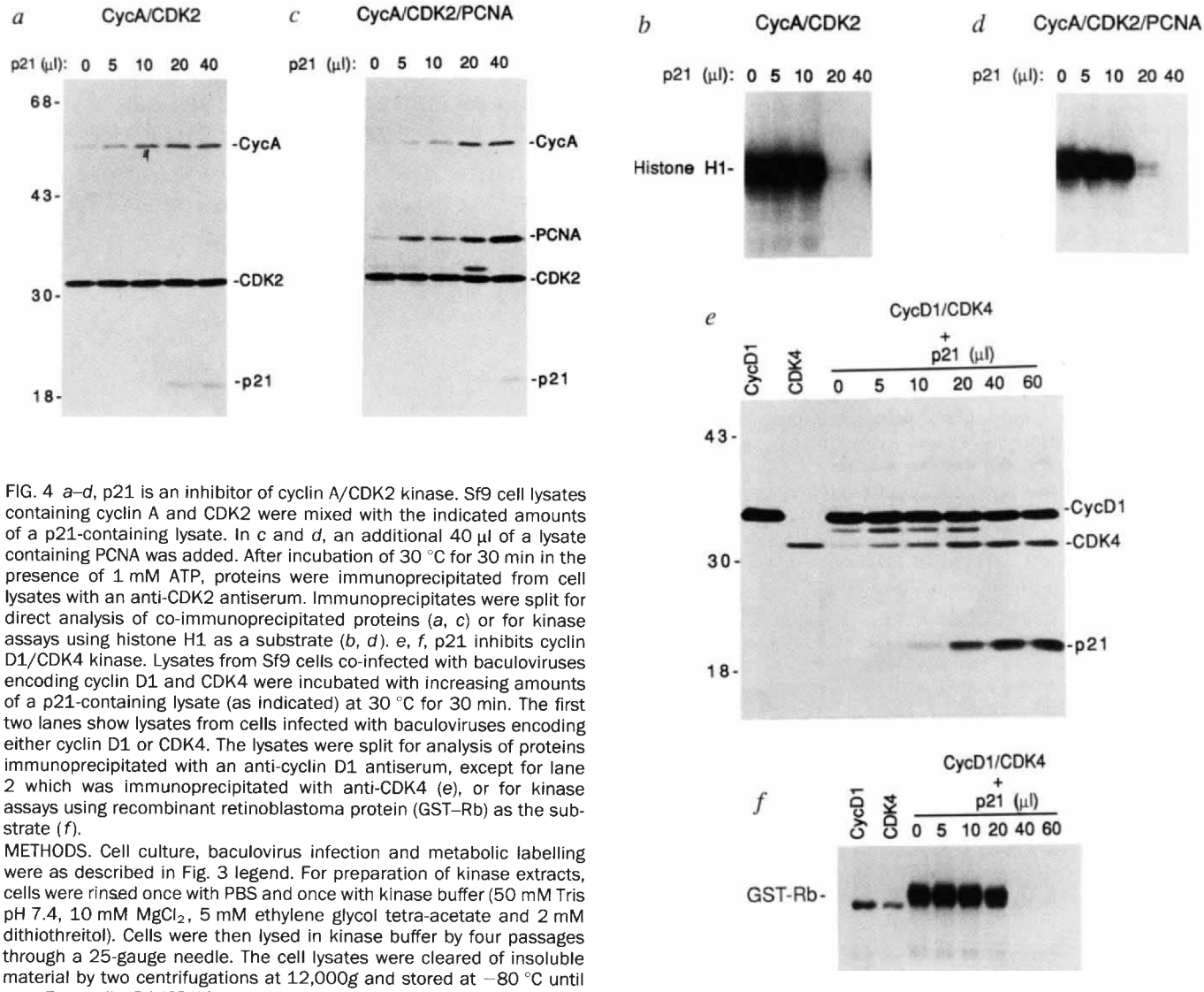


FIG. 4 *a–d*, p21 is an inhibitor of cyclin A/CDK2 kinase. Sf9 cell lysates containing cyclin A and CDK2 were mixed with the indicated amounts of a p21-containing lysate. In *c* and *d*, an additional 40 μl of a lysate containing PCNA was added. After incubation of 30 °C for 30 min in the presence of 1 mM ATP, proteins were immunoprecipitated from cell lysates with an anti-CDK2 antiserum. Immunoprecipitates were split for direct analysis of co-immunoprecipitated proteins (*a*, *c*) or for kinase assays using histone H1 as a substrate (*b*, *d*). *e*, *f*, p21 inhibits cyclin D1/CDK4 kinase. Lysates from Sf9 cells co-infected with baculoviruses encoding cyclin D1 and CDK4 were incubated with increasing amounts of a p21-containing lysate (as indicated) at 30 °C for 30 min. The first two lanes show lysates from cells infected with baculoviruses encoding either cyclin D1 or CDK4. The lysates were split for analysis of proteins immunoprecipitated with an anti-cyclin D1 antiserum, except for lane 2 which was immunoprecipitated with anti-CDK4 (*e*), or for kinase assays using recombinant retinoblastoma protein (GST–Rb) as the substrate (*f*).

METHODS. Cell culture, baculovirus infection and metabolic labelling were as described in Fig. 3 legend. For preparation of kinase extracts, cells were rinsed once with PBS and once with kinase buffer (50 mM Tris pH 7.4, 10 mM $MgCl_2$, 5 mM ethylene glycol tetra-acetate and 2 mM dithiothreitol). Cells were then lysed in kinase buffer by four passages through a 25-gauge needle. The cell lysates were cleared of insoluble material by two centrifugations at 12,000*g* and stored at −80 °C until use. For cyclin D1/CDK4 complexes, the lysates were from cells co-infected with baculoviruses encoding cyclin D1 and CDK4. For *a–d*, lysates containing cyclin A, CDK2, p21 or PCNA were mixed in the presence of 1 mM ATP for 30 min at 30 °C. In all experiments, differences in reaction volumes were compensated for with lysates from Sf9 cells infected with wild-type baculovirus. After incubation, the reactions were stopped by adding 20 mM EDTA and the cyclin/CDK protein complexes were immunoprecipitated as described[11]. After immunoprecipitation, the samples were split in half; half was used for the analysis of the protein composition and half for kinase assay. Histone H1 kinase assays were performed as described[12]. For the analysis of cyclin D1/CDK4 kinase activity, lysates were incubated at 30 °C for 30 min without ATP and kinase activity was assayed using bacterially produced GST–Rb as described[13].

However, cyclin D/CDK4 binary complexes catalysed substantial Rb phosphorylation (Fig. 4*e*, *f*). Addition of increasing amounts of p21 resulted in the accumulation of cyclin D/CDK4/p21 ternary complexes (Fig. 4*e*), with a corresponding inhibition of Rb phosphorylation (Fig. 4*f*). Again, inclusion of PCNA was essentially without effect (our unpublished results).

The ability of p21 to inhibit such disparate cyclin/CDK kinases suggests that p21 is a universal CDK inhibitor. Thus, overexpression of p21 *in vivo* would be expected to cause cell-cycle arrest. To address this question, we examined the effect of p21 overexpression on cell proliferation using a stable colony formation assay[7]. Transfection of SAOS-2 cells with a pRcCMV vector alone yielded a large number of stable transformants. However, transfection with either of two independent preparations of a plasmid directing the overexpression of p21 failed to produce an appreciable number of colonies (results not shown). The effect of p21 overexpression was virtually identical to the effect of p53 overexpression in a parallel transfection. These results indicate that p21 may be an inhibitor of cell proliferation.

As we have previously found that p21 is absent from cyclin/CDK complexes in cells lacking functional p53 (ref. 1), we isolated the murine p21 cDNA (data not shown) and examined p21 messenger RNA levels in fibroblasts derived from p53-'null' mice. Compared with fibroblasts from normal embryos, p53 'null' fibroblasts showed ~50-fold lower levels of p21 mRNA (data not shown). Furthermore, p21 mRNA is induced ~10-fold by γ-irradiation of a $p53^+$ myeloid leukaemia cell line (mL-7) but is unchanged upon similar treatment of a myeloid leukaemia cell line that lacks p53 (HL-60; data not shown). These results indicate that p21 is regulated by the p53 pathway.

In many transformed cells, cyclins and CDKs associate in binary complexes which form the core of the cell-cycle regulatory machinery. In normal cells, a major fraction of the cyclin kinases acquires two additional subunits and thereby forms quaternary complexes[4,5]. We have isolated a cDNA encoding the one uncharacterized component of these quaternary complexes, p21. Reconstitution of quaternary complexes in insect cells revealed that p21 is a universal inhibitor of cyclin kinases. As such, p21 inhibits cell proliferation upon overexpression in mammalian cells. Taken in conjunction with the previously demonstrated absence of p21 protein in the cell-cycle kinase complexes of cells with deficient p53, our results indicate that p21 could be a transcriptional target of the tumour suppressor protein, p53. One function of p53 is to act in a cell signalling pathway which causes cell-cycle arrest following DNA damage (see, for example ref. 8). We suggest that p21 forms a critical link between p53 and the cell-cycle control machinery. □

Received 12 November; accepted 1 December 1993.

1. Xiong, Y., Zhang, H. & Beach, D. *Genes Dev.* **7,** 1572–1583 (1993).
2. Sherr, C. *Cell* **73,** 1059–1065 (1993).
3. Draetta, G. *Trends biochem. Sci.* **15,** 378–383 (1990).
4. Xiong, Y., Zhang, H. & Beach, D. *Cell* **71,** 505–514 (1992).
5. Zhang, H., Xiong, Y. & Beach, D. *Molec. Biol. Cell* **4,** 897–906 (1993).
6. Matsushime, H. *et al. Cell* **71,** 323–334 (1992).
7. Zhu, L. *et al. Genes Dev.* **7,** 1111–1125 (1993).
8. Kastan, M. B. *et al. Cell* **71,** 587–597 (1992).
9. Kawasaki, H., Emori, Y. & Suzuki, K. *Analyt. Biochem.* **191,** 332–336 (1990).
10. Xiong, Y., Menninger, J., Beach, D. & Ward, D. *Genomics* **13,** 575–584 (1992).
11. Hannon, G. J., Demetrick, D. & Beach, D. *Genes Dev.* **7,** 2378–2391 (1993).
12. Draetta, G. & Beach, D. *Cell* **54,** 17–26 (1988).
13. Ewen, M. E. *et al. Cell* **73,** 487–497 (1993).

ACKNOWLEDGEMENTS. We thank J. Smith and P. Burfiend for assistance with preparation of the manuscript and flow cytometry, respectively, M. Ockler, P. Renna and J. Duffy for help with photography and graphics, M. Nichols, S. Matsumoto and C. Gawel for technical assistance with cloning and sequencing the p21 cDNA, S. Allan for help with the colony formation suppression assays, and M. Serrano for discussion and for help in compiling figures. Y.X. thanks B. Marzluff for his help and encouragement. We also acknowledge the gift of baculovirus reagents from D. Morgan (UCSF) and S. Gruenwald (Pharmingen). Y.X. is supported by the program in molecular biology of the University of North Carolina, Chapel Hill. D.B. is an investigator of the Howard Hughes Medical Institute. H.Z. is a Research Associate of the Howard Hughes Medical Institute. G.J.H. is a postdoctoral fellow of the Damon Runyon-Walter Winchell Cancer Research Fund. This work was supported by the NIH. Y.X. and G.H. contributed equally to this work.

A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4

Manuel Serrano, Gregory J. Hannon & David Beach*

Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, New York 11724, USA

THE division cycle of eukaryotic cells is regulated by a family of protein kinases known as the cyclin-dependent kinases (CDKs)[1,2]. The sequential activation of individual members of this family and their consequent phosphorylation of critical substrates promotes orderly progression through the cell cycle[3,4]. The complexes formed by CDK4 and the D-type cyclins have been strongly implicated in the control of cell proliferation during the G1 phase[3–6]. CDK4 exists, in part, as a multi-protein complex with a D-type cyclin, proliferating cell nuclear antigen and a protein, p21 (refs 7–9). CDK4 associates separately with a protein of M_r 16K, particularly in cells lacking a functional retinoblastoma protein[9]. Here we report the isolation of a human p16 complementary DNA and demonstrate that p16 binds to CDK4 and inhibits the catalytic activity of the CDK4/cyclin D enzymes. p16 seems to act in a regulatory feedback circuit with CDK4, D-type cyclins and retinoblastoma protein.

The yeast two-hybrid protein interaction screen[10] was used to search for proteins that can associate with human CDK4. Two-hybrid screening relies on reconstituting a functional GAL4 transcriptional activator from two separate fusion proteins, the activation domain ($GAL4^{ad}$) and the DNA-binding domain ($GAL4^{db}$). A positive cDNA clone was found which contained, in-phase with $GAL4^{ad}$, an open reading frame of 148 amino acids encoding a protein of M_r 15,845 comprising four ankyrin repeats (Fig. 1*a*). We have named this protein $p16^{INK4}$ (inhibitor of CDK4; see below).

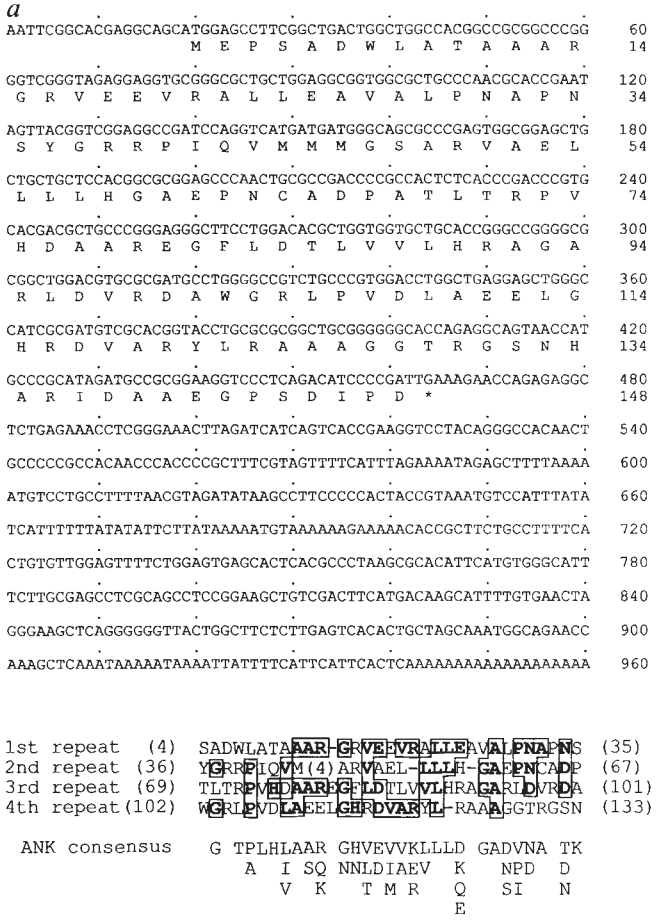
To test the specificity of the association between $p16^{INK4}$ and CDK4, yeast cells were co-transformed with a plasmid encoding the $GAL4^{ad}$–$p16^{INK4}$ fusion and with plasmids encoding several different targets (see Fig. 1*b*). Only the $GAL4^{db}$–CDK4 fusion interacted with GAL^{ad}–$p16^{INK4}$ to an extent that allowed growth in the absence of histidine (Fig. 1*b*). The specificity of this interaction was studied in a cell-free system. A fusion protein consisting of glutathione *S*-transferase fused to $p16^{INK4}$ (GST–$p16^{INK4}$) was expressed in bacteria and purified. GST–$p16^{INK4}$ was mixed with different *in vitro*-translated ^{35}S-labelled CDKs (Fig. 1*c*, top), and the GST–$p16^{INK4}$ fusion protein was recovered from the different mixtures on glutathione-Sepharose beads. GST–$p16^{INK4}$ bound much more efficiently to CDK4 (>30-fold) than to the other CDKs tested (Fig. 1*c*, middle). The specificity of the CDK4/$p16^{INK4}$ interaction was also studied in insect cells infected with a recombinant baculovirus encoding $p16^{INK4}$ and with baculoviruses encoding CDK4 or CDK2, respectively (Fig. 1*d*). $p16^{INK4}$ was co-immunoprecipitated with anti-CDK4 (Fig. 1*d*, lane 1), but not with anti-CDK2 (lane 4). Conversely $p16^{INK4}$ antibodies co-immunoprecipitated CDK4 (Fig. 1*d*, lane 3), but not CDK2 (lane 6). These results demonstrate that $p16^{INK4}$ interacts specifically with CDK4. Glycerol gradient centrifugation indicated that CDK4 and $p16^{INK4}$ from insect cell extracts form a binary (1:1) complex (data not shown).

Anti-CDK4 immunoprecipitates from a normal human diploid fibroblast line, W138, revealed that CDK4 associates with several proteins, cyclin D1, proliferating cell nuclear antigen (PCNA), p21 and p16 (Fig. 2*a*, lane 1). Proteins present in this immunoprecipitate probably represent at least two independent

* To whom correspondence should be addressed.

complexes, one comprising the quaternary complex of CDK4, cyclin D1, p21 and PCNA, and a binary complex containing CDK4 and p16 (refs 7–9). In contrast, in VA13 and HeLa cells, CDK4 is predominantly, if not exclusively, associated with p16 (Fig. 2*a*, lanes 5 and 9, respectively[9]). VA13 is a SV40 virus-transformed derivative of W138, and HeLa cells express the papillomavirus E6 and E7 proteins. Anti-p16^{INK4} immunoprecipitates contained a protein of M_r 16K which was readily detectable in the transformed cell lines VA13 and HeLa (Fig. 2*a*, lanes 7 and 11, respectively) and to a much lesser extent in the normal cell line W138 (lane 3). We also found that at least two other proteins, p33 and p38, specifically co-immunoprecipitate with p16^{INK4} (Fig. 2*a*, lanes 3, 7, 11). The *N*-chlorosuccinimide (NCS) partial proteolytic pattern of the CDK4-associated p16 was identical to the pattern generated from immunoprecipitated p16^{INK4} (Fig. 2*b*, compare lanes 1 and 2). In addition, the partial V8 protease patterns of the p16-associated p33 and CDK4 were identical (Fig. 2*b*, compare lanes 3 and 4 with lanes 5 and 6, respectively). These results show unequivocally that our cDNA encodes the CDK4-associated p16. The other apparent p16^{INK4}-associated protein, p38, might correspond to the CDK4-related kinase PSLIRE whose M_r is close to 38K (ref. 11).

We have reconstituted active CDK4/cyclin D complexes that can phosphorylate a fusion protein consisting of gluthathione *S*-transferase and a fragment of the retinoblastoma protein (Rb) termed the large pocket[12,13] (Fig. 3*a*, lanes 1, 2). Addition of extracts containing p16^{INK4} abolished phosphorylation of GST–Rb by cyclin D2/CDK4 (Fig. 3*a*, lanes 3–5) whereas extracts



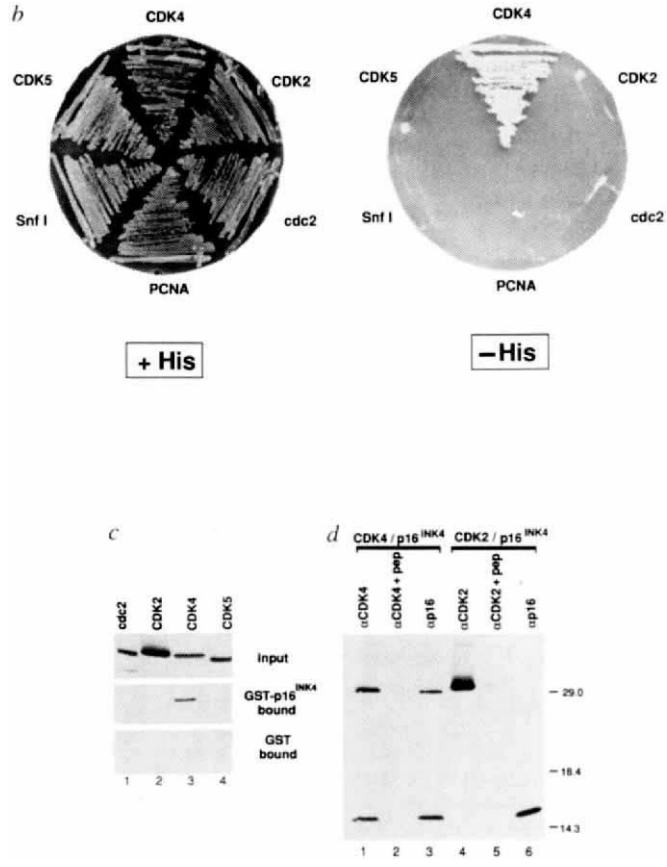


FIG. 1 Sequence of the human p16^{INK4} cDNA and specific interaction between p16^{INK4} and CDK4. *a*, Top, sequence of p16^{INK4}. We have noted a motif at the amino terminus of p16^{INK4} ($MX_3ADWLATAX_4RVEEVX_2LL$) which shows homology to the amino-terminus of the cyclin box[19]. This is the same region, termed the P-box, which in B-type cyclins contributes to the activation of the cdc25 phosphatase[20]. However, mutation of this motif affected neither CDK4 binding nor inhibition of CDK4 kinase by p16^{INK4} (our unpublished results). Bottom, p16^{INK4} is formed by four ankyrin repeats[21]. *b*, Yeast cells were simultaneously transformed with a plasmid expressing a GAL4^{db}–INK4 fusion and with plasmids expressing the GAL4^{ad} fused to the indicated CDK, PCNA or the budding yeast kinase, Snfl. Cells containing both plasmids were streaked on plates with or without histidine. The ability to grow in the absence of histidine depends on the expression of the *HIS3* gene that is under the control of a GAL4-responsive promoter. *c*, Purified, bacterially produced GST–p16^{INK4} fusion protein was mixed with ^{35}S-labelled *in vitro*-translated CDKs, as indicated. Top, analysis of an aliquot of the *in vitro*-translation products. Centre, ^{35}S-labelled proteins recovered on glutathione–Sepharose beads following incubation with GST–p16^{INK4}; 1% of the input CDK4 was recovered in lane 3. Bottom, ^{35}S-labelled proteins recovered following incubation with GST. Purification of GST–p16^{INK4} and binding assays were done as described[15]. *d*, ^{35}S-labelled insect cells lysates expressing CDK4 and p16^{INK4} (lanes 1–3) or CDK2 and p16^{INK4} (lanes 4–6) were immunoprecipitated with the indicated antibodies.

METHODS. *a*, For two-hybrid screening, human CDK4 (ref. 6) was fused to the GAL4 DNA-binding domain in the pGBT9 vector (P. Bartel and S. Fields, unpublished data). The HeLa cDNA library was constructed as described[15] except that a low-expression derivative (pGAD-GL) of pGAD-GH was used which lacks a 900-base pair (bp) *Sph*I fragment from the alcohol dehydrogenase promoter. Screening was done as described[15]. *d*, The recombinant baculovirus expressing p16^{INK4} was constructed using the vector pVL1393 and the Baculo-Gold kit (Pharmingen). Baculoviruses expressing CDK2 and CDK4 were obtained from D. Morgan and H. Zang, respectively. Preparation of cell lysates and immunoprecipitation were as described[15]. Antisera against GST–p16^{INK4} were generated in rabbits by Pocono Rabbit Farm and Laboratory, Inc. Antibodies were immunoaffinity purified using a CNBr-activated Sepharose 4B column (Pharmacia) as described[22]. The anti-CDK4 peptide antibody has been described previously[9]. The anti-CDK2 antibody was raised against a synthetic peptide corresponding to the carboxy-terminal region of human CDK2 (provided by K. Galaktionov).

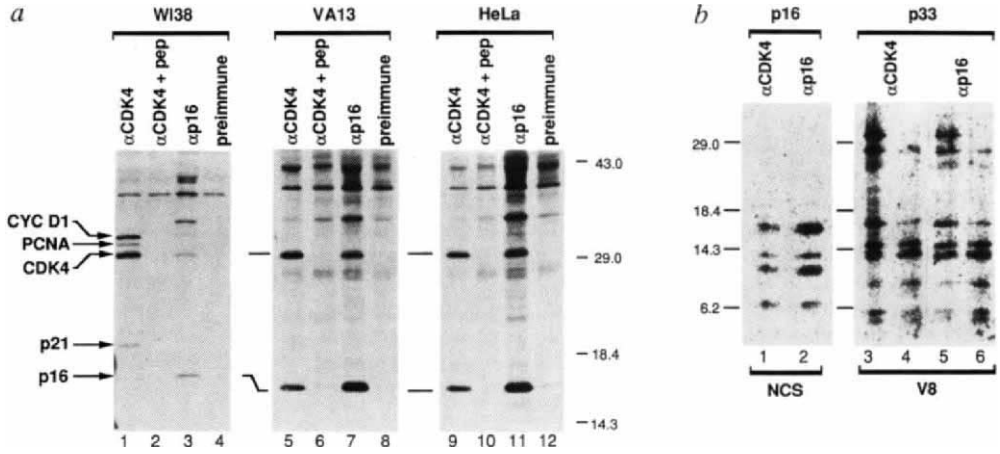


FIG. 2 p16^{INK4} associates with CDK4 in human cells. *a*, Proteins were immunoprecipitated from ^{35}S-labelled lysates of W138 (lanes 1–4), VA13 (lanes 5–8) or HeLa (lanes 9–12) cells with the indicated antibody. The identity of the proteins marked with arrows was determined previously for the anti-CDK4 immunoprecipitates[7,9], and in this report for the anti-p16^{INK4} immunoprecipitates (see *b*). *b*, p16 co-immunoprecipitated from HeLa cell lysates with anti-CDK4 (lane 1) and p16^{INK4} immunoprecipitated from HeLa cell lysates with anti-p16^{INK4} (lane 2) were digested with 15 mM NCS and digestion products were analysed by SDS–PAGE. CDK4 immunoprecipitated from HeLa cell lysates (lanes 3, 4) was compared with the p16^{INK4}-associated p33 (lanes 5, 6) by partial digestion with V8 protease, 100 ng (lanes 3, 5) or 500 ng (lanes 4, 6).

METHODS. *a*, Immunoprecipitations were done as described[7]. The immunoprecipitates were separated in a 12% polyacrylamide gel (30:1 acrylamide:bis-acrylamide). In this gel system, the mobility of PCNA with respect to cyclin D1 is reversed (data not shown) in comparison with the gel system used in other reports (12.5% polyacrylamide, 125:1 acrylamide:bis-acrylamide[7–9,18]). *b*, In-gel partial digestion with V8 protease was performed as described[22]. For NCS partial digestion, gel slices were rehydrated by soaking in water and equilibrated in NCS buffer (8.3 M urea; 50% acetic acid). NCS was added and gel slices were incubated for 30 min at room temperature. Digestion was stopped by soaking in water. Before electrophoresis, gel slices were equilibrated in gel buffer (62.5 mM Tris–HCl pH.7.0, 10% glycerol, 3% SDS, 15% β-mercaptoethanol). Electrophoresis was done in gels containing 17.5% polyacrylamide (250:1 acrylamide:bis-acrylamide).

containing p53 (lanes 6–8) had no effect. Identical results were obtained using extracts containing CDK4/cyclin D1 and CDK4/cyclin D3 (data not shown). We have also found that two other members of the Rb protein family, p107 and p130 (refs 14–16), are phosphorylated by cyclin D/CDK4 kinase *in vitro*, and that p16^{INK4} inhibits phosphorylation of these substrates (data not shown). The CDK2/cyclin D2 kinase also phosphorylates GST–Rb *in vitro*[13]. Neither p16^{INK4} (Fig. 3*a*, lanes 11–13) nor p53 (lanes 14–16) inhibited the CDK2/cyclin D2 kinase, suggesting that the inhibition of CDK4 might be related to the interaction between p16^{INK4} and CDK4. To confirm that this inhibition was due solely to the presence of p16^{INK4} in the insect cell lysates, we purified a histidine-tagged p16^{INK4} protein produced in insect cells (Fig. 3*b*). The purified His–p16^{INK4} protein inhibited the activity of the CDK4/cyclin D2 complex (Fig. 3*c*, lanes 3–6) whereas a mock purification from

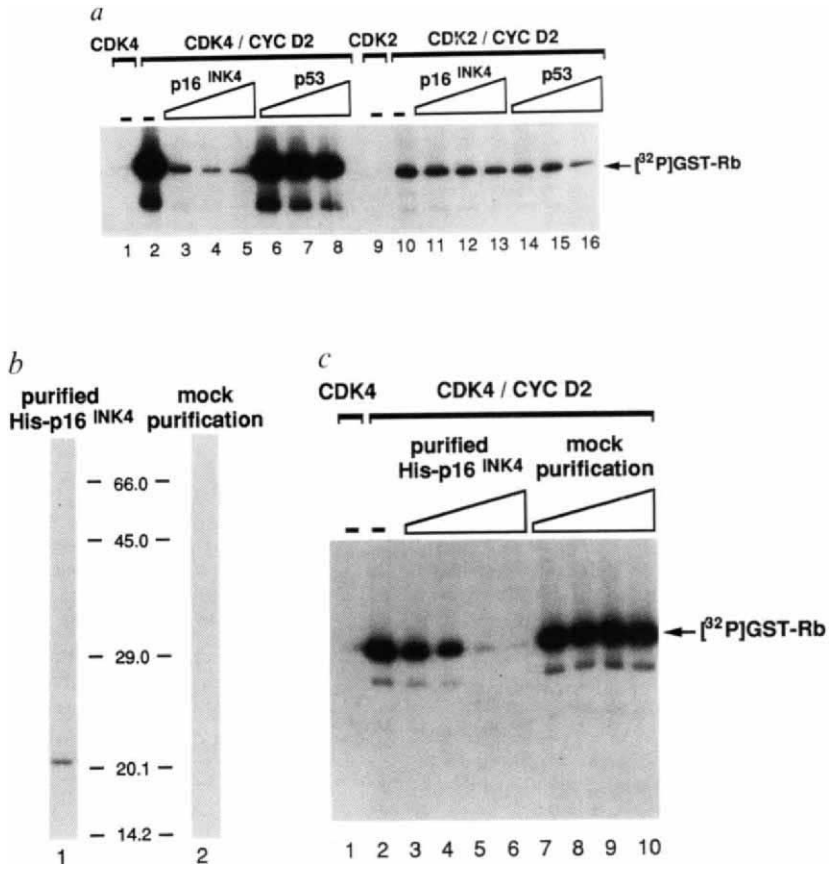


FIG. 3 p16^{INK4} inhibits CDK4/cyclin D2 kinase. *a*, Extracts (2 μl) from baculovirus-infected insect cells expressing CDK4 (lane 1), CDK4 plus cyclin D2 (lanes 2–8), CDK2 (lane 9) or CDK2 plus cyclin D2 (lanes 10–16) were mixed with different amounts (1 μl, lanes 3, 6, 11, 14; 2 μl, lanes 4, 7, 12, 15; and 4 μl, lanes 5, 8, 13, 16) of similar extracts containing p16^{INK4} (lanes 3–5, 11–13) or human p53 (lanes 6–8, 14–16). Mixtures were then assayed for their ability to phosphorylate GST–Rb. *b*, Coomassie-blue-stained gel with purified His-p16^{INK4} (lane 1) and an equivalent volume of the mock purification (lane 2). *c*, Extracts (2 μl) containing CDK4 (lane 1) or CDK4 plus cyclin D2 (lanes 2–10) were mixed with increasing amounts of purified His-p16^{INK4} (final concentrations: 2.5 ng μl^{-1}, lane 3; 5 ng μl^{-1}, lane 4; 10 ng μl^{-1}, lane 5; 20 ng μl^{-1}, lane 6) or with the equivalent volume of the mock purification, as indicated in the figure (lanes 7–10), and tested for their ability to phosphorylate GST–Rb.

METHODS. *a*, The recombinant baculovirus directing the expression of cyclin D2 was obtained from S. Gruenwald. Extracts were prepared as described[12]. For kinase assays, extracts were mixed in kinase buffer (50 mM Tris–HCl pH 7.5, 10 mM $MgCl_2$, 1 mM dithiothreitol) up to a final volume of 10 μl, and incubated for 10 min at 30 °C. Reactions were started by addition of 2 μl containing 125 ng of purified GST–Rb large-pocket protein[12,13] (the plasmid was provided by M. Ewen), 25 μM ATP and 5 μCi[^{32}P-γ]ATP (3,000 Ci mmol; NEN). Reactions proceeded for 1 min at 30 °C, and were stopped by addition of 50 μl of a solution 0.5% NP40 and glutathione–Sepharose beads. Beads were loaded onto a 10% polyacrylamide gel. *b*, A recombinant baculovirus expressing His-tagged p16^{INK4} was constructed using vector pAcSG-His-NT-C and the Baculo-Gold kit (Pharmingen). Purification was done as described[23].

insect cells infected with a non-recombinant baculovirus (expressing the polyhedrin protein) produced no detectable inhibition (Fig. 3*c*, lanes 7–10).

The biochemical properties of $p16^{INK4}$ suggest that it could act as a negative regulator of the proliferation of normal cells. The relative abundance of D-type cyclins and $p16^{INK4}$ could determine the activity of the CDK4 kinase and thus regulate cell-cycle progression. The exclusive presence of the p16/CDK4 complex in cells transformed by a variety of DNA tumour viruses poses a paradox. This can, however, be resolved assuming that Rb and related 'pocket proteins' are the exclusive critical substrates of cyclin D/CDK4. If this is the case, in those cells lacking functional Rb the activity of CDK4 would be unnecessary and thus the inhibitory role of $p16^{INK4}$ would be without effect. Consistent with this, it has been found that microinjection of cyclin D1 antibodies into SV40-transformed fibroblasts or Rb-defective SAOS-2 cells fails to cause cell-cycle arrest[17]. $p16^{INK4}$ may act in normal cells in a negative feedback loop whose role is to downregulate CDK4 once Rb has been inactivated by phosphorylation. In cells in which Rb is constitutively inactive, $p16^{INK4}$ expression is increased, with consequent inhibition of CDK4. However, in this situation, the negative feedback loop is futile as CDK4 kinase is not required for cell-cycle progression.

We have identified two different negative regulators of cyclin kinases. Our data suggest that p21 affects cell-cycle arrest induced by p53 (ref. 18). $p16^{INK4}$ is proposed to act both upstream and downstream of Rb to form a negative feedback loop which regulates the ability of Rb to prevent cell proliferation. □

Received 12 November; accepted 1 December 1993.

1. Solomon, M., J. *Curr. Opin. Cell Biol.* **5,** 180–186 (1993).
2. Nigg, E. A. *Curr. Opin. Cell Biol.* **5,** 187–193 (1993).
3. Motokura, T. & Arnold, A. *Biochim. biophys. Acta* **1155,** 63–78 (1993).
4. Sherr, C. J. *Cell* **73,** 1059–1065 (1993).
5. Matsushime, H., Roussel, M. F., Ashmun, R. A. & Sherr, C. J. *Cell* **65,** 701–713 (1991).
6. Matsushime, H. *et al. Cell* **71,** 323–334 (1992).
7. Xiong, Y., Shang, H. & Beach, D. *Cell* **71,** 505–514 (1992).
8. Zhang, H., Xiong, Y. & Beach, D. *Molec. biol. Cell* **4,** 897–906 (1993).
9. Xiong, Y., Zhang, H. & Beach, D. *Genes Dev.* **7,** 1572–1583 (1993).
10. Fields, S. & Song, O. *Nature* **340,** 245–246 (1989).
11. Meyerson, M. *et al. EMBO J.* **11,** 2909–2917 (1992).
12. Kato, J.-Y., Matsushime, H., Hiebert, S. W., Ewen, M. E. & Sherr, C. J. *Genes Dev.* **7,** 331–342 (1993).
13. Ewen, M. E. *et al. Cell* **73,** 487–497 (1993).
14. Zhu, L. *et al. Genes Dev.* **7,** 1111-1125 (1993).
15. Hannon, G. J., Demetrick, D. & Beach, D. *Genes Dev.* **7,** 2378–2391 (1993).
16. Li, Y., Graham, C., Lacy, S., Duncan, A. M. V. & Whyte, P. *Genes Dev.* **7,** 2366–2377 (1993).
17. Lukas, J., Pagano, M., Staskova, Z., Draetta, G. & Bartek, J. *Oncogene* (in the press).
18. Xiong, Y. *et al. Nature* **366,** 701–704 (1993).
19. Lees, E. M. & Harlow, E. *Molec. cell. Biol.* **13,** 1194–1201 (1993).
20. Zheng, X.-F. & Ruderman, J. V. *Cell* **75,** 155–164 (1993).
21. Michaely, P. & Bennet, V. *Trends Cell Biol.* **2,** 127–129 (1992).
22. Harlow, E. & Lane, D. *Antibodies: A Laboratory Manual* (Cold Spring Habor Laboratory Press, New York, 1988).
23. Chen, X.-S., Brash, A. R. & Funk, C. D. *Eur. J. Biochem.* **214,** 845–852 (1993).

ACKNOWLEDGEMENTS. We thank P. Bartel and S. Fields (SUNY, Stony Brook, New York) for providing plasmids and strains before publication, D. Morgan (UCSF), S. Gruenwald (Pharmingen) and C. Prives (Columbia University) for giving us different recombinant baculoviruses, M. Ewen (Dana Farber Cancer Institute) for providing the plasmid that directs the expression of GST–Rb, G. Draetta and M. Pagano (Mitotix) for communicating their unpublished results, members of our laboratory, H. Zhang, K. Galaktionov and D. Casso for critical reading of the manuscript, and C. Gawel and S. Matsumoto for technical assistance. M.S. is recipient of a long-term EMBO fellowship, and G.J.H. is a postdoctoral fellow of the Damon Runyon–Walter Winchell Cancer Research Fund. D.B. is an investigator of the Howard Hughes Medical Institute. This work was supported in part by the NIH.

Inhibition of CDK2 activity *in vivo* by an associated 20K regulatory subunit

Yong Gu*, Christoph W. Turck† & David O. Morgan*‡

Departments of * Physiology and † Medicine, and † Howard Hughes Medical Institute, University of California, San Francisco, California 94143, USA

THE major events of the cell division cycle are triggered by periodic changes in the activity of cyclin-dependent protein kinases (CDKs). In mammals, the members of the CDK family include CDK2 and CDC2, which are thought to be involved in the control of DNA replication and mitosis, respectively[1–3]. The protein kinase activity of these enzymes is controlled by a complex array of mechanisms[4–6]. Activation of the CDK catalytic subunit requires association with a positive regulatory subunit (cyclin) and phosphorylation (at Thr 160 in CDK2). This activated complex can be inhibited by additional phosphorylation at Thr 14 and Tyr 15. Here we report the identification of a new mechanism for the regulation of CDK2 activity. We find that CDK2/cyclin complexes in mouse fibroblasts associate tightly with a 20K protein (CAP20). Complexes containing CAP20 were isolated from cell lysates and found to have negligible kinase activity, indicating that CAP20 association *in vivo* may inhibit CDK2 activity. We purified CAP20 from 3T3 cells and found that low concentrations of the protein completely inhibit the kinase activity of CDK2 *in vitro*. Thus CAP20 represents a new negative regulatory subunit that inhibits the activity of CDK2/cyclin complexes in mammalian cells.

‡ To whom correspondence should be addressed.

Our search for new regulators of CDK2 began with analyses of proteins that associate with CDK2 in mouse BALB/c 3T3 fibroblasts. When CDK2 is immunoprecipitated from metabolically labelled lysates of these cells, two other proteins are also specifically immunoprecipitated (Fig. 1*a*, lanes 1–4). The larger protein ($M_r \sim 55{,}000$ (55K)) is recognized by antibodies directed against cyclin A (results not shown) and probably represents cyclin A, the major cyclin partner for CDK2. The nature of the smaller protein ($M_r \sim 20$K) is not known; we refer to this protein as CDK2-associated protein-20 (CAP20). CAP20 is not detectable in immunoprecipitates of CDC2 (Fig. 1*a*, lanes 5 and 6).

We found that cation-exchange chromatography can be used to isolate a subpopulation of CDK2 that is associated with CAP20. Metabolically labelled cell lysates were passed over an S-Sepharose column, which was washed and then eluted with low salt (100 mM NaCl) and then with high salt (500 mM NaCl). Immunoprecipitation analysis revealed that CDK2 in the high-salt eluate, but not in other fractions, is associated with cyclin A and CAP20 (Fig. 1*b*). Immunoprecipitates of cyclin A from the high-salt eluate also contained CAP20 (Fig. 1*c*). Thus CAP20 is present in immunoprecipitates of both CDK2 and cyclin A from this fraction, suggesting that CAP20 is associated with CDK2/cyclin A complexes.

We next used a similar procedure to separate CDK2/cyclin complexes containing CAP20 from those lacking the protein. Crude cell lysates were injected onto a Mono-S column, which was washed with 100 mM NaCl and then eluted with a linear salt gradient up to 500 mM NaCl. CDK2 eluted at two distinct positions: peak I eluted at ~250 mM NaCl and Peak II eluted at ~450 mM NaCl (Fig. 2*a*, *b*). Gel filtration analysis showed that CDK2 in both peaks migrated at the large size characteristic of CDK/cyclin complexes (Fig. 2*c*). Interestingly, measurement of CDK2 activity in Mono-S fractions revealed that CDK2 in peak II was essentially inactive (Fig. 2*b*). We estimate that the

specific kinase activity of CDK2 in peak I is over 30-fold greater than that in peak II.

To determine whether CAP20 was associated with either of these CDK2 subpopulations, CDK2/cyclin complexes from pooled peaks I or II were purified by immunoaffinity chromatography and analysed by polyacrylamide gel electrophoresis and silver staining. CDK2 in peak II, but not peak I, was found to be associated with CAP20 (Fig. 2*d*). Additional proteins present only in peak II were not seen reproducibly in our experiments with metabolically labelled cells (Fig. 1) and may represent nonspecifically bound proteins.

To explain the inactivity of the CDK2 in peak II, we first explored previously known mechanisms by which CDK2 may be inhibited. First, the CDK2 in peak II is not simply lacking a cyclin partner: gel filtration indicates that CDK2 in peak II migrates in a high-molecular-weight complex (Fig. 2*c*), and cyclin A copurifies with CDK2 during immunoaffinity chromatography (Fig. 2*d*). Second, CDK2 in peak II does not lack phosphorylation at the activating site Thr 160. Immunoblotting reveals that CDK2 in these fractions migrates as a doublet, and our previous work[7] has shown that the lower band represents CDK2 that is phosphorylated at Thr 160 (Fig. 2*a*, *c*). Third, CDK2 in peak II is not inhibited by phosphorylation at Thr 14 or Tyr 15, because specific dephosphorylation of these sites with the phosphatase CDC25 does not increase CDK2 activity[7] (not shown). Finally, treatment with purified cyclin A and partially purified Thr 160 kinase (CDK-activating kinase or CAK) does not activate the CDK2 in peak II (not shown).

Thus, the CDK2/cyclin complexes in peak II are inhibited by a novel mechanism. Our observation that CAP20 is associated with these complexes suggests that CAP20 may play a part in this mechanism. To investigate this possibility, we purified CAP20 and tested its effect on CDK2 activity *in vitro*. Large amounts of CDK2/cyclin/CAP20 complexes were purified from 3T3 cells by S-Sepharose and immunoaffinity chromatography and subjected to SDS-polyacrylamide gel electrophoresis. A gel fragment containing CAP20 was excised, and the protein was extracted from the fragment and renatured (Fig. 3).

We tested the effect of renatured CAP20 on the activity of several human CDK/cyclin complexes. Baculovirus-derived CDK and cyclin proteins were mixed *in vitro*, immunoprecipitated, and activated by phosphorylation with partially purified human CAK. Immunoprecipitates were incubated with CAP20 for 30 min, washed, and analysed for protein kinase activity (Fig.

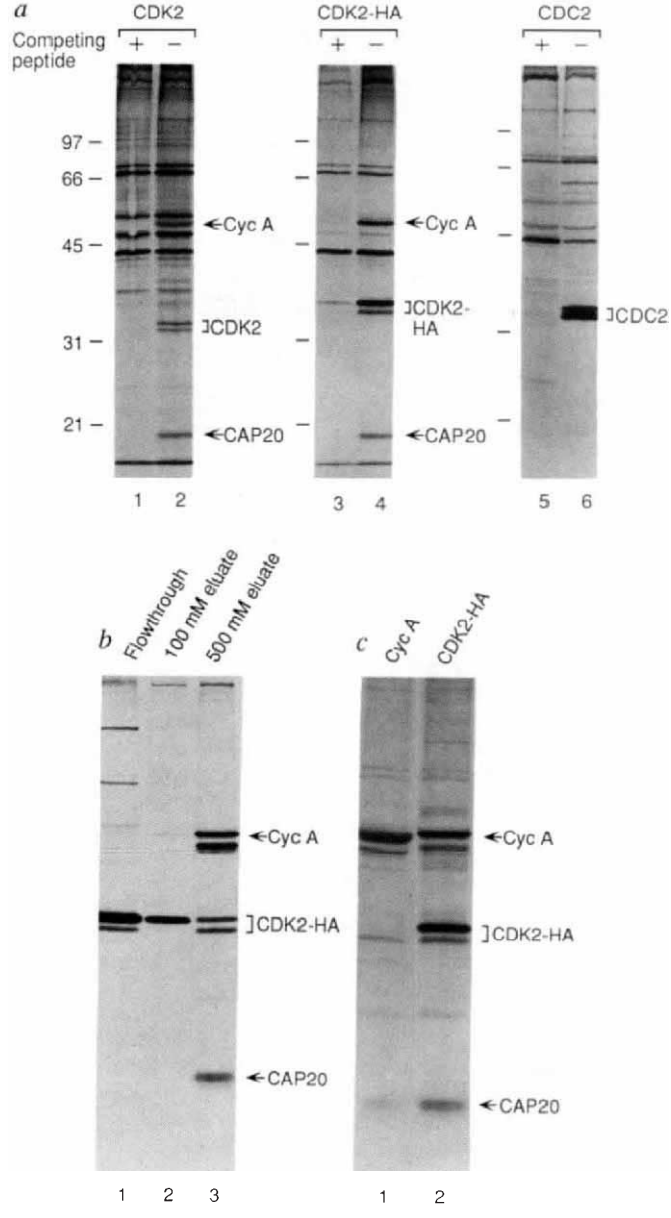


FIG. 1 Association of CDK2/cyclin complexes with a 20K protein in 3T3 mouse fibroblasts. *a*, Lysates were prepared from ^{35}S-labelled 3T3 cells and immunoprecipitated with affinity-purified antibodies against the C terminus of CDK2 (lanes 1 and 2) or CDC2 (lanes 5 and 6). Lysates were also prepared from a stably transfected 3T3 cell line constitutively expressing a version of human CDK2 containing a C-terminal haemagglutinin (HA) epitope tag (CDK2-HA), and subjected to immunoprecipitation with a monoclonal antibody (mAb 12CA5) directed against the HA tag (lanes 3 and 4). Immunoprecipitations were done in the presence (lanes 1, 3 and 5) or absence (lanes 2, 4 and 6) of excess antigenic peptides. Immunoprecipitated proteins were separated on 11% polyacrylamide gels and visualized by autoradiography. *b*, ^{35}S-labelled lysates from CDK2-HA cells were fractionated on S-Sepharose and immunoprecipitated with mAb 12CA5, followed by electrophoresis in a 12% polyacrylamide gel and autoradiography. *c*, A metabolically labelled high-salt eluate from an S-Sepharose column (as in *b*) was immunoprecipitated with anti-cyclin A antiserum (lane 1). For comparison, lane 2 shows a 12CA5 immunoprecipitate of an ^{35}S-labelled lysate of CDK2-HA cells.

METHODS. Mouse BALB/c 3T3 fibroblasts (clone A31) and the CDK2-HA stable transfectant were cultured as described[7]. Cells were metabolically labelled by 6 h growth in methionine/cysteine-free DME-H16 medium supplemented with 10% dialysed calf serum, 50 μg ml^{-1} L-methionine, 10 mM HEPES, pH 7.4, and 0.1 mCi ml^{-1} ^{35}S-methionine/cysteine (Express protein labelling mix, New England Nuclear). Lysates were prepared as described[7] and immunoprecipitated with affinity-purified anti-CDK2 or anti-CDC2 antibodies, or monoclonal antibody 12CA5, as described[7]. Polyclonal anti-cyclin A serum was a gift from J. Pines. To fractionate small amounts of ^{35}S-labelled cell lysates by cation-exchange chromatography, a 350-μl S-Sepharose fast-flow column was prepared and pre-equilibrated with buffer A (10 mM HEPES, pH 7.4, 10 mM NaCl, 1 mM EDTA). Metabolically labelled lysates (~3 mg) were diluted fivefold in buffer A and passed over the column several times. The column was washed with buffer A, and bound proteins were eluted first with 1 ml of buffer A containing 100 mM NaCl (100 mM eluate) and then with 1 ml buffer A containing 500 mM NaCl (500 mM eluate).

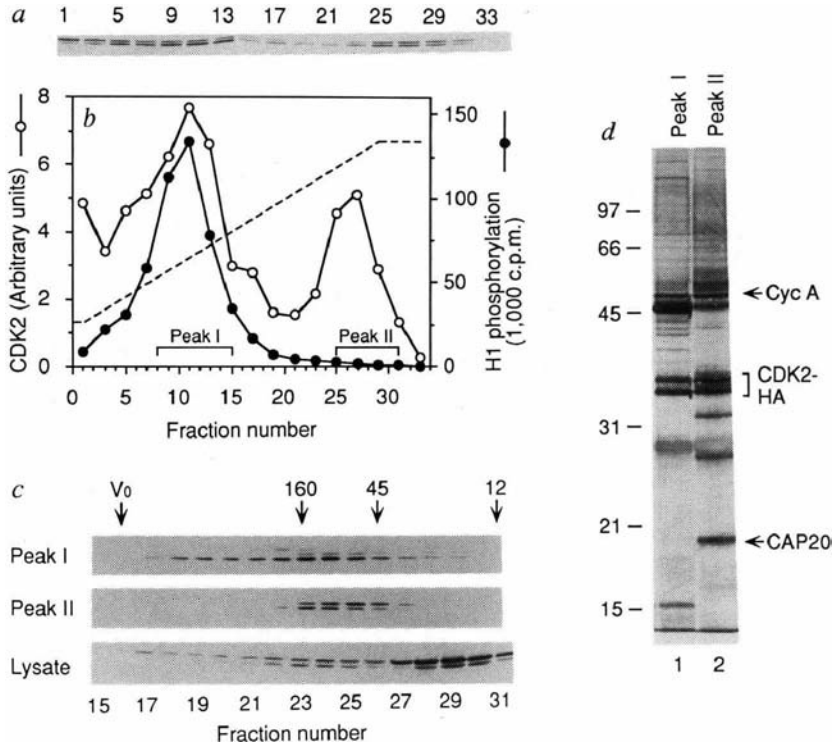


FIG. 2 Association of CAP20 with an inactive subpopulation of CDK2/cyclin complexes. *a*, CDK2-HA cell lysate was injected onto a Mono-S column, which was washed and then eluted with a linear salt gradient (100–500 mM NaCl). Aliquots of every second fraction were electrophoresed on an 11% polyacrylamide gel and immunoblotted with mAb 12CA5 to visualize the CDK2-HA protein. Numbers represent fraction numbers. *b*, CDK2 protein in Mono-S fractions was quantified by incubating the 12CA5 immunoblot with ^{125}I-labelled anti-mouse antibodies, followed by quantification with a PhosphorImager[17] (open circles). Histone H1 kinase activity was also measured in 12CA5 immunoprecipitates of each fraction (filled circles). The dashed line indicates the linear salt gradient. *c*, Peaks I and II from the Mono-S fractions in *b* were individually pooled, concentrated and fractionated on a Superose-12 gel filtration column. CDK2-HA protein in each fraction was visualized by immunoblotting with mAb 12CA5. For comparison, the bottom panel shows gel filtration of a crude CDK2-HA cell lysate. Note that the third band above the CDK2 doublet in peak I fractions is nonspecific. Molecular weights (in thousands) of marker proteins, determined in parallel runs, are indicated (V_0 = void volume; IgG, 160; ovalbumin, 45; cytochrome *c*, 12). *d*, Pooled Mono-S fractions from peak I and peak II were loaded on an immunoaffinity column consisting of mAb 12CA5 covalently linked to Protein A–Agarose. After washing, bound proteins were eluted with an alkaline buffer and visualized by silver staining after electrophoresis on a 12% polyacrylamide gel.

METHODS. Cell lysates were prepared from twenty-five 15-cm dishes of confluent CDK2-HA cells (~150 mg total protein in 8 ml), diluted fivefold with buffer A supplemented with protease inhibitors (Fig. 1) and injected at 0.5 ml min^{-1} onto a Pharmacia Mono-S HR5/5 column pre-equilibrated with buffer A. The column was washed with buffer A containing 100 mM NaCl and then eluted with a 30-ml linear gradient (100–500 mM NaCl in buffer A) at 0.5 ml min^{-1}. 1 ml fractions were collected. Immunoblotting with 12CA5 and measurement of histone H1 kinase activity in 12CA5 immunoprecipitates were performed as described[7]. Gel filtration was done as described[17]. Immunoaffinity chromatography with mAb 12CA5 is described in Fig. 4.

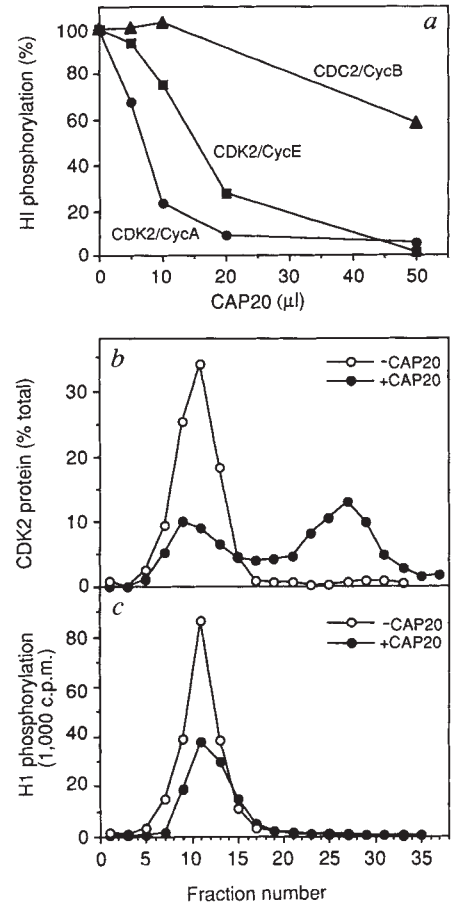


FIG. 3 Effect of purified CAP20 on CDK2 activity and chromatographic behaviour. *a*, The indicated amounts of renatured CAP20 protein were incubated 30 min with immunoprecipitates containing the indicated CDK/cyclin complexes. After washing, histone H1 kinase activity was measured in immunoprecipitates and is expressed as a percentage of activity in buffer-treated control immunoprecipitates. CDK2 activity in these experiments was not affected by the addition of a control renatured protein (see Methods). *b*, Fraction 10 from Mono S peak I (Fig. 2*b*) was incubated with 1 ml CAP20 (filled circles) or control buffer (open circles) and re-chromatographed on the Mono-S column as described for Fig. 2. CDK2 in every second fraction was quantified by immunoblotting. *c*, CDK2-HA-associated histone H1 kinase activity was determined in immunoprecipitates of Mono-S fractions from the experiment in *b*.

METHODS. Immunoaffinity-purified CDK2/cyclin/CAP20 complexes, prepared as described in Fig. 4, were mixed with 25 µg β-lactoglobulin A, denatured in SDS and separated on an 11% polyacrylamide gel. A control lane contained only β-lactoglobulin. The gel was stained briefly with Coomassie blue, and 3-mm slices containing CAP20 or the control protein were excised. Proteins were extracted overnight at 24 °C by shaking crushed gel slices in 1 ml extraction buffer[18] (50 mM Tris-HCl, pH 8.0, 0.1 mM EDTA, 200 mM NaCl, 1 mM DTT, 0.1% SDS, 100 µg ml^{-1} BSA) and precipitated with 4 vols cold (−20 °C) acetone. Proteins were dissolved in 20 µl 6 M guanidine-HCl in dilution buffer (25 mM HEPES, pH 7.4, 200 mM NaCl, 1 mM EDTA, 10% glycerol, 1 mM DTT). After 20 min at 24 °C, reactions were diluted 50-fold with dilution buffer and incubated for another 20 min. We estimate that the CAP20 concentration in these preparations is 0.5–1 µg ml^{-1} (25–50 nM). To obtain the indicated CDK/cyclin complexes, baculovirus-infected insect cell lysates containing CDKs or cyclins were mixed as described[17,19], immunoprecipitated with mAb 12CA5, and activated with partially purified human CAK (R. Fisher, unpublished results). Immunoprecipitates, containing ~10 ng CDK per sample, were washed and then incubated 30 min at 24 °C with the indicated amount of CAP20 or renatured β-lactoglobulin (not shown) in a total volume of 50 µl.

```
Mouse CAP20   Q T S L T D F Y H S K
Human p21     Q T S M T D F Y H S K

Mouse CAP20   V Y L S P G S R
Human p21     L Y L P T G P R

Mouse CAP20   S L G L P K
Human p21     G L G L P K
```

FIG. 4 Amino-acid sequences from mouse CAP20. The amino-acid sequences of three peptides obtained from purified mouse CAP20 protein are aligned with the corresponding sequences from human p21 (ref. 20). Of the 25 amino acid residues compared, 19 (boxed) residues (76%) are identical.

METHODS. To purify large amounts of CAP20 protein, lysates (~2.5 g total protein) were prepared from CDK2-HA cells (350 15-cm dishes in a typical preparation) and diluted fivefold in buffer A (Fig. 2) plus protease inhibitors. Diluted lysates were clarified by centrifugation (100,000g for 30 min) and loaded at 1 ml min^{-1} onto an S-Sepharose fast-flow column (1.5 cm × 15 cm; 27 ml) pre-equilibrated with buffer A. After washing with buffer A plus 100 mM NaCl, bound proteins were eluted with 50 ml of buffer A containing 500 mM NaCl. Eluted proteins (~250 mg) were then passed twice over a period of 2 h through an immunoaffinity column containing 5 mg mAb 12CA5 covalently linked to 1 ml protein A–Agarose (prepared using the Affinica antibody orientation kit; Schleicher & Schuell). The column was then washed with 20 ml wash buffer (50 mM Tris-HCl, pH 7.4, 250 mM NaCl, 5 mM EDTA, 1 mM EGTA, 50 mM NaF, 0.1% TritonX-100, 0.05% SDS). Bound proteins were eluted with freshly prepared 100 mM triethylamine (pH 11.5) and collected in 0.5 ml fractions. Fractions containing CAP20 (revealed by SDS–PAGE and silver staining) were pooled and brought to 10% trichloroacetic acid. Precipitated proteins were pelleted, dissolved in SDS–PAGE sample buffer and analysed on 12% polyacrylamide gels. Proteins were electroblotted onto PVDF membrane (Immobilon-P, Millipore, Bedford, MA) in CAPS buffer (10 mM CAPS pH 11.0, 10% methanol). Membrane containing CAP20 protein, visualized by Ponceau S staining, was excised and incubated with 1.5% polyvinylpyrrolidone (PVP-40; Sigma) in methanol to block protein binding sites. CAP20 protein (typically 5 μg per preparation) was digested with 1 μg sequencing-grade trypsin at 37 °C for 24 h and the resulting tryptic peptides were subjected to reverse-phase HPLC on a narrow-bore column. Three peptides were sequenced by Edman degradation with an Applied Biosystems gas-phase sequencer (Model 475A).

3*a*). Low concentrations of CAP20 protein (in the 10–20 nM range) completely inhibited the kinase activity of CDK2/cyclin A or CDK2/cyclin E complexes, whereas activity was unaffected by samples from control gel slices lacking protein or containing a control protein (β-lactoglobulin) (not shown). Under these conditions, inhibition of CDC2/cyclin B complexes required 5–10-fold higher concentrations of CAP20. This is consistent with our observation that CAP20 is not detectable in CDC2 immunoprecipitates from 3T3 cells.

These experiments with purified CAP20 *in vitro* suggest that the inactivity of the CDK2/cyclin complexes in Mono-S peak II is due to the presence of CAP20. To explore this possibility, a peak I fraction containing active CDK2 was incubated with CAP20, and the resulting mixture was re-chromatographed on Mono-S. When we added the amount of CAP20 needed to inhibit a significant fraction of the CDK2 population, a similar fraction of CDK2 protein no longer eluted at the position of peak I but now eluted at the position of peak II (Fig. 3*b*). The shifted protein had negligible kinase activity (Fig. 3*c*). Thus the association of CAP20 with CDK2/cyclin complexes abolishes their kinase activity and alters their chromatographic behaviour on the Mono-S column. This suggests that the fraction of CDK2 in peak II may be a useful measure of CAP20-inhibited CDK2 in the cell. About 30% of CDK2/cyclin complexes are found in peak II, indicating that CAP20 may be suppressing the activity of a significant fraction of the CDK2/cyclin complexes in the cell.

We were unable to do the converse experiment of removing CAP20 from the complexes in peak II and determining whether activity was restored. Treatment of CDK2/cyclin/CAP20 complexes with a variety of chaotropic salts and detergents indicated that CAP20 does not dissociate except under conditions that dissociate the entire complex.

To characterize CAP20 at the molecular level, we obtained amino-acid sequence from purified mouse CAP20 protein. The sequences of three tryptic peptides from CAP20 were determined by Edman degradation (Fig. 4). These sequences indicate that CAP20 is probably a mouse homologue of human p21, whose cDNA was recently cloned[20]. The p21 protein has been found in immunoprecipitates of various CDKs (CDK2, CDK4 and CDC2) and cyclins (cyclins A, B and D1) from several human cell lines[8,9]. This protein also appears to be identical to a human protein (SDI1), recently identified as the product of a senescent cell complementary DNA[21]. In addition, others have isolated the same cDNA with a yeast screen for human cDNAs encoding proteins that interact with CDK2 (ref. 22).

The existence of negative regulatory subunits for mammalian CDK proteins adds another level of complexity to an enzyme system that is already very tightly controlled. The function of this new regulatory protein in cell cycle control remains unclear, as we have found that the amount of CAP20 associated with CDK2 does not appear to change dramatically during the normal cell cycle (Y.G. and D.O.M., unpublished results). It is possible that CAP20 is regulated post-translationally, or may simply serve to inhibit CDK2/cyclin complexes in specific subcellular locations. CAP20 may also provide a mechanism for negative growth control under certain specialized conditions.

CDK2, in association with various cyclins, has been implicated in the control of Start and the initiation of DNA replication[10–16]. CAP20, as a potent negative regulator of CDK2, is in a position to have major control over these processes. For this reason the loss of CAP20 from the cell might be expected to impair growth control and perhaps contribute to cell transformation and neoplasia. □

Received 25 October; accepted 19 November 1993.

1. Norbury, C. & Nurse, P. *A. Rev. Biochem.* **61,** 441–470 (1992).
2. Reed, S. I. *A. Rev. Cell Biol.* **8,** 529–561 (1992).
3. Pines, J. *Trends biochem. Sci.* **18,** 195–197 (1993).
4. Solomon, M. *Curr. Opin. Cell Biol.* **5,** 180–186 (1993).
5. De Bondt, H. L. *et al. Nature* **363,** 595–602 (1993).
6. Draetta, G. *Trends Cell Biol.* **3,** 287–289 (1993).
7. Gu, Y., Rosenblatt, J. & Morgan, D. O. *EMBO J.* **11,** 3995–4005 (1992).
8. Xiong, Y., Zhang, H. & Beach, D. *Cell* **71,** 505–514 (1992).
9. Xiong, Y., Zhang, H. & Beach, D. *Genes Dev.* **7,** 1572–1583 (1993).
10. Fang, F. & Newport, J. W. *Cell* **66,** 731–742 (1991).
11. Pagano, M., Pepperkok, R., Verde, F., Ansorge, W. & Draetta, G. *EMBO J.* **11,** 961–971 (1992).
12. Pagano, M. *et al. J. Cell Biol.* **121,** 101–111 (1993).
13. Girard, F., Strausfeld, U., Fernandez, A. & Lamb, N. J. C. *Cell* **67,** 1169–1179 (1991).
14. Ohtsubo, M. & Roberts, J. M. *Science* **259,** 1908–1912 (1993).
15. Sherr, C. J. *Cell* **73,** 1059–1065 (1993).
16. Tsai, L.-H., Lees, E., Faha, B., Harlow, E. & Riabowol, K. *Oncogene* **8,** 1593–1602 (1993).
17. Desai, D., Gu, Y. & Morgan, D. O. *Molec. Biol. Cell* **3,** 571–582 (1992).
18. Hager, D. A. & Burgess, R. R. *Analyt. Biochem.* **109,** 76–86 (1980).
19. Koff, A. *et al. Science* **257,** 1689–1694 (1992).
20. Xiong, Y. *et al. Nature* **366,** 701–704 (1993).
21. Noda, A., Ning, Y., Venable, S. F., Pereira-Smith, O. M. & Smith, J. R. *Expl Cell Res.* (in the press).
22. Harper, J. W., Adami, G. R., Wei, N., Keyomarsi, K. & Elledge, S. J. *Cell* **75,** 805–816 (1993).

ACKNOWLEDGEMENTS. We thank H. Jones for insect cell lysates, R. Fisher for partially purified human CAK, A. Murray for advice, R. Fisher and S. Murphy for comments on the manuscript, and Y. Xiong, D. Beach, J. Smith, S. Elledge and W. Harper for sharing unpublished data. Y.G. is supported by a California Division-American Cancer Society senior fellowship. This work was supported by grants (to D.O.M.) from the NIH, the Markey Charitable Trust, the March of Dimes Birth Defects Foundation and the Rita Allen Foundation.